Investigation of DNA Repair Pathways and the Impact of SMC Protein on Chromosome Stability in Mycobacteria

---

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Carolin Güthlein

aus

Deutschland

Promotionskomitee
Prof. Dr. Erik C. Böttger (Vorsitz)
Dr. Burkhard Springer (Leitung der Dissertation)
Prof. Dr. Josef Jiricny

Zürich 2008

ACKNOWLEDGEMENTS

This thesis would not have been possible without the important support of many people.

First of all, I would like to express my gratitude to my supervisors Prof. Dr. Erik C. Böttger and Dr. Burkhard Springer who inspired and supported this work. I am grateful for the possibility to work in their laboratory and deeply appreciate their detailed and constructive comments in uncountable fruitful discussions.

Sincere thanks to P.D. Dr. Peter Sander for revising the publication drafts and for providing helpful suggestions.

I would also like to thank Prof. Dr. Josef Jiricny for his agreement to co-examine this work and our collaboration partner Dr. Elaine O. Davis (National Institute for Medical Research, London, UK) for performing the mouse infection experiments.

Many thanks to all of the present and past lab members for the pleasant atmosphere in the laboratory, and for their assistance in all aspects of everyday work. Special thanks go to the members of the Springer Group: Dr. Roger Wanner for assisting me in statistics and design of the graphs, Dr. Silke Peter for support and helpful discussions, and Martin Bosshard for technical assistance in performing the experiments.

Sincere thanks to all my friends, especially to Dr. Ben R. Townsend for offering me support and cheering me up in good and in bad times during the writing of this thesis.

Finally, thanks to my family, especially to my parents. Their encouragement and understanding importantly contributed to the achievement of all of my aims.

Thank you.

CONTENTS

Summary

Annually, *Mycobacterium tuberculosis* is responsible for nearly two million deaths worldwide. Part of its success lies in its ability to survive and replicate in macrophages, but the molecular mechanisms allowing it to do so are not yet understood. Once activated, macrophages produce reactive oxygen (ROI) and nitrogen (RNI) species, which amongst other targets damage DNA. *M. tuberculosis* seems to be superbly adapted to these unfavourable conditions, which demand efficient mechanisms to ensure genome stability.

Mycobacteria are naturally devoid of the highly conserved mismatch repair (MMR) system, a crucial mechanism for mutation avoidance. Other DNA repair strategies, such as recombinational repair, base excision repair (BER), or nucleotide excision repair (NER), may compensate for some of the functions accomplished by the MMR system. In this study, the role of the NER pathway in mycobacteria was investigated. The bacterial NER pathway consists of the UvrABC endonuclease enzyme complex and a helicase named UvrD. Knockout mutants in the excinuclease component (*uvrA* or *uvrB,* respectively) and in the helicase (*uvrD*) were constructed in *M. tuberculosis* and in *M. smegmatis*. In addition, double mutants lacking both components were generated. Analyses of the NER mutants revealed that the mycobacterial NER system repairs a wide range of mutagenic DNA alterations and is an important defence mechanism against oxidative and nitrosative damage. In a gene conversion assay, the capability of *M. smegmatis* NER mutants to recognise and subsequently reject different base-pairing errors was assessed. Inactivation of *uvrB* and *uvrD* increased marker integration frequencies to various extents, with in part – and dependent on the mismatch studied - synergistic effects in the combined *uvrB* / *uvrD* mutant. Our results imply that NER and particularly the helicase UvrD in part compensate for the lack of MMR in mycobacteria. We also found that the helicase component UvrD is essential for long-term survival of *M. tuberculosis in vitro* and *in vivo*.

Besides DNA repair, other mechanisms contribute to ensuring genome integrity. DNA binding proteins that establish chromosome architecture presumably are involved in stabilising the bacterial genome during persistence and to accomplish proper cell division during reactivation. SMC (structural maintenance of chromosomes) proteins play fundamental roles in various aspects of chromosome organisation and dynamics, including repair of DNA damage. Mutant strains of *M. smegmatis* and *M. tuberculosis* defective in SMC were

constructed. Surprisingly, inactivation of *smc* did not result in a recognizable phenotype. This is in contrast to data on *smc* null mutants in other species. Our observations suggest that the maintenance of mycobacterial chromosome organisation differs from that of described models.

Taken together, our observations provide evidence that despite the lack of a MMR system, mycobacteria possess efficient defence mechanisms against DNA damage. We found that the NER system has a fundamental role in mutation avoidance and that the helicase UvrD has an additional role besides participation in NER. In addition, our data suggest unique mechanisms for chromosome organisation and partitioning in mycobacteria. Overall, the obtained findings indicate that mechanisms that govern genome stability in mycobacteria differ significantly from those of described model species.

Zusammenfassung

Jedes Jahr ist das Bakterium *Mycobacterium tuberculosis* für weltweit fast zwei Millionen Todesfälle verantwortlich. Diese Bilanz liegt zum Teil darin begründet, dass das Bakterium in Makrophagen überleben und sich replizieren kann, wobei die molekularen Mechanismen für diese Fähigkeit noch nicht aufgeklärt sind. Aktivierte Makrophagen bilden reaktive Sauerstoff- (engl. reactive oxygen species, ROI) und Stickstoffmoleküle (engl. reactive nitrogen species, RNI), die unter anderem DNA-Schäden hervorrufen. *Mycobacterium tuberculosis* scheint hervorragend an diese Bedingungen angepasst zu sein, welche effiziente Mechanismen zur Sicherung der Genomstabilität erfordern.

Mykobakterien fehlt das äusserst konservierte Basenfehlpaarungsreparatursystem (engl. mismatch repair, MMR), welches ein wichtiger Mechanismus zur Vermeidung von Mutationen ist. Es wird vermutet, dass andere DNA-Reparaturstrategien, wie Reparatur durch homologe Rekombination, Basenexzisionsreparatur (engl. base excision repair, BER) oder Nukleotidexzisionsreparatur (engl. nucleotide excision repair, NER) manche Funktionen des MMR übernehmen könnten. In dieser Arbeit wurde die Funktion des NER in Mykobakterien untersucht. Der bakterielle Nukleotidexzisionsreparaturweg besteht aus dem UvrABC Endonukleaseenzymkomplex und der Helikase UvrD. Die Exzinukleasekomponente *uvrA* bzw. *uvrB* sowie die Helikase *uvrD* von *Mycobacterium tuberculosis* und *Mycobacterium smegmatis* wurden durch gezielten Genaustausch inaktiviert. Zusätzlich wurden Doppelmutanten generiert, denen beide Komponenten fehlen.
Die Charakterisierung der NER Mutanten zeigte, dass das mykobakterielle NER-System viele Arten von mutagenen DNA-Veränderungen reparieren kann und einen wichtigen Abwehrmechanismus gegen oxidative und nitrosative DNA-Schäden darstellt. In einem Genkonversionsassay wurde untersucht, ob die *M. smegmatis* NER-Mutanten verschiedene Basenfehlpaarungen erkennen und aussondern können. Verlust von *uvrB* und *uvrD* führte zu signifikant erhöhten Genkonversionsfrequenzen, sowie teilweise – abhängig von der untersuchten Basenfehlpaarung – zu einem synergistischen Effekt bei der *uvrB / uvrD* Doppelmutante. Unsere Ergebnisse zeigen, dass das NER- System und besonders die Helikase UvrD dazu beitragen, das Fehlen des MMR-Systems in Mykobakterien zu kompensieren. Ausserdem konnten wir *in vitro* und *in vivo* zeigen, dass die Helikase UvrD essentiell für die Persistenz von *M. tuberculosis* ist.

Neben der DNA-Reparatur sind weitere Mechanismen an der Gewährleistung der Genomintegrität beteiligt. Es wird angenommen, dass Chromosomenstrukturproteine wesentlich dazu beitragen, dass das Genom während der Persistenz stabilisiert wird und dass die Zellteilung während der Reaktivierung korrekt verläuft. SMC (structural maintenance of chromosomes) Proteine haben grundlegende Funktionen in vielen Aspekten der Chromosomenorganisation und Chromosomendynamik. *Mycobacterium smegmatis* und *Mycobacterium tuberculosis smc* Mutanten wurden durch gezielten Genaustausch generiert. Erstaunlicherweise prägten die *smc* Mutanten keinen ersichtlichen Phänotyp aus. Dieses Ergebnis steht im Gegensatz zu bisher beschriebenen *smc* Mutanten in anderen Organismen. Unsere Beobachtungen deuten darauf hin, dass die mykobakterielle Chromosomorganisation von beschriebenen Modellen abweicht.

Die Ergebnisse dieser Arbeit lassen sich wie folgt zusammenfassen. Obwohl Mykobakterien das MMR-System fehlt, verfügen sie über effiziente Abwehrmechanismen zur Vermeidung von DNA-Schäden. Das NER-System hat eine bedeutende Funktion in der Vermeidung von Mutationen und die Helikase UvrD hat neben der Beteiligung im NER-System weitere Funktionen. Des Weiteren weisen unsere Beobachtungen auf aussergewöhnliche Mechanismen der Chromosomenorganisation und Chromosomenteilung in Mykobakterien hin. Die Resultate zeigen, dass die Mechanismen zur Erhaltung der Genomstabilität in Mykobakterien signifikant von denen in bisher beschriebenen Modellorganismen abweichen.

Abbreviations

BER	base excision repair
bp	basepair
CFU	colony forming units
kbp	Kilobasepairs
kDa	Kilodalton
KO	knockout
NER	nucleotide excision repair
NHEJ	nonhomologous end-joining
MBC	minimal bactericidal concentration
MDR-TB	multidrug –resistant *M. tuberculosis*
MIC	minimal inhibitory concentration
MMC	mitomycin C
MMR	mismatch repair
OD	optical density
ORF	open reading frame
RNI	reactive nitrogen intermediates
ROI	reactive oxygen intermediates
sco	single cross-over
SMC	structural maintenance of chromosomes
TB	tuberculosis
TBH	*tert*-butylhydroperoxide
UV-C	short wavelength UV light
wt	wildtype
XDR-TB	extensively drug-resistant *M. tuberculosis*

1 Introduction

1.1 *Mycobacterium tuberculosis* and disease

With more than eight million new cases each year, tuberculosis (TB) remains a global health problem. The World Health Organisation (WHO) estimates that approximately 1.6 million deaths per year are tuberculosis-related. More than 90 percent of TB cases and 98 percent of TB deaths occur in the developing world (data from "Tuberculosis facts- Handout 2007" published on www.who.int). TB is a chronic, contagious disease, which is caused by the members of the *Mycobacterium tuberculosis* complex.

The *Mycobacterium tuberculosis* complex includes the closely related organisms *M. tuberculosis*, *M. africanum*, *M. bovis*, *M. bovis BCG, M. caprae, M. pinnipedii, M. microti* and *M. canettii* (Brosch *et al.*, 2002). Despite their close genetic similarity, these organisms differ significantly with regard to epidemiology, pathogenicity, and host spectrum. *M. tuberculosis* is considered the principal cause of TB in humans. Most people infected with *M. tuberculosis* contain the initial infection and develop latent, asymptomatic TB. Following primary infection, *M. tuberculosis* can persist for decades within the human body and is able to cause active disease at a later date. It has been estimated that one-third of the world's population, amounting to around 2 billion people, is latently infected (Dye *et al.*, 2005).

In spite of the fact that potent antitubercular drugs are available, the course of therapy requires a treatment period of at least six months and is rather complex in cases of drug resistant TB (Böttger, 2006; Furin, 2007; Shah *et al.*, 2007). The long duration of treatment is thought to result from bacteria shifting into a non-replicating or dormant state in the host which renders the bacteria tolerant to antibiotic killing (Wayne and Hayes, 1996). HIV infection, which weakens the immune system and allows reactivation of latent TB infection, as well as the emergence of multidrug-resistant TB (MDR-TB), pose a serious threat to TB control (Böttger and Springer, 2008). Resistance to anti-TB drugs primarily arises due to poorly managed TB care. As a result of incorrect drug prescribing practices, poor quality drugs and patient non-adherence, extensively drug-resistant (XDR) *M. tuberculosis* strains have evolved during the last years (Shah *et al.*, 2007). As XDR *M. tuberculosis* strains are resistant to the most powerful first-line and second-line drugs, patients are left with treatment options that are less effective resulting in unfortunate treatment outcomes. The emergence of multidrug resistance in *M. tuberculosis* constitutes an alarming trend. Consistent with the genetic isolation of this organism, acquisition of novel transmissible genetic elements by horizontal gene transfer is extremely rare, if found at all (Fleischmann *et al.*, 2002; Hirsh *et al.*, 2004). In fact, all known

resistance mechanisms of *M. tuberculosis* are due to spontaneous mutations in chromosomal genes (Böttger, 1994; Finken *et al.*, 1993; Meier *et al.*, 1996; Ramaswamy and Musser, 1998). One key factor to pathogenesis is the ability of *M. tuberculosis* to persist albeit the host immune response. Following initial infection, the bacteria survive and replicate within macrophages, normally part of the host's defence system (Kaufmann, 2001). The bacilli are exposed to changing host environments, e.g. nutritional deprivation, hypoxia, reactive nitrogen and oxygen species. It is assumed that the bacteria are exposed to nitric oxygen during latent infection, as inhibition of inducible nitric oxide synthase in mouse models of latency leads to reactivation of bacterial replication (Flynn *et al.*, 1998). Nitric oxygen generates reactive species capable of damaging bacterial cell components. Reactive oxygen intermediates (ROI) and reactive nitrogen intermediates (RNI) can damage lipids, proteins and DNA (Chan *et al.*, 1992).

Elevated mutation rates can promote adaptation to fluctuating living conditions. Mutator strains are characterised by an increased spontaneous mutation rate due to defects in genes involved in DNA repair or error avoidance systems. While a mutator phenotype may be a driving force in adaptive evolution, intact DNA repair is essential during persistent infection. If not repaired, DNA damage in a haploid microorganism may cause extinction, either immediately, by blocking chromosome replication, or on a longer time scale, by accumulation of unfavourable mutations. Hence, the survival of species depends on a balance between genome stability and the requirement for mutation to allow adaptation to novel environmental conditions. In general, the selective advantage of a high mutation rate is transient and maintenance of genomic integrity is essential for long-term survival of the population (Warner and Mizrahi, 2006). Thus, it can be reasoned that DNA repair mechanisms play a crucial role in long-term survival of *M. tuberculosis* in the host.

1.2 Role of DNA repair in genome stability

1.2.1 Common principles in DNA repair

DNA repair mechanisms are well conserved from bacteria to humans, due to their importance in maintaining genome integrity. All living organisms are equipped with DNA repair systems that can cope with a wide variety of DNA lesions. The proteins involved in DNA repair can be grouped into the following major functional categories: damage reversal, excision repair and repair of DNA double-strand breaks (Friedberg, 2005).

Damage reversal

Direct repair systems reverse the mutagenic event. Repair by direct reversal mechanisms are extremely specific to the type of damage incurred (Mishina *et al.*, 2006; Sancar, 2003). To circumvent the necessity of individual repair pathways for each of the numerous types of DNA alterations, cells are endowed with more general repair systems (Friedberg, 2005).

Excision repair

Multiple distinct mechanisms exist for excising damaged parts of the DNA, termed nucleotide excision repair (NER), base excision repair (BER) and mismatch repair (MMR). The principle of all three mechanisms involves splicing out the damaged region and inserting new bases to fill the gap (Friedberg, 2005).

The nucleotide excision repair system displays a broad substrate specifity: NER recognises all lesions that generate bulky distortions in the conformation of the DNA double helix, e.g. UV-induced photolesions (e.g. pyrimidine dimers), intrastrand-crosslinks and large chemical adducts in the DNA formed through exposure to genotoxic agents (Batty and Wood, 2000). Eukaryotes and prokaryotes share the basic mechanism although the proteins involved are different (Batty and Wood, 2000). The process of nucleotide excision repair in bacteria is mediated by the UvrABC endonuclease enzyme complex and a helicase named UvrD. The process of NER begins with the binding of a complex of an UvrA dimer and one UvrB protein to the DNA, scanning for damage in an ATP-dependent manner. On encountering damage, the UvrA dimer dissociates and UvrC is recruited. The UvrBC complex holds an excinuclease activity and cuts the DNA several bases upstream and downstream of the damage. After replacing UvrC, UvrD removes the excised DNA strand with the altered base. Subsequently, the resulting gap is filled in by polymerase and the nick is sealed by ligase I (Sancar, 1996).
A common type of DNA damage is base damage, which is usually corrected by BER, involving the action of DNA-glycosylases. These enzymes specifically recognise damaged bases and cleave the glycosidic bond between the damaged base and the deoxyribose backbone. Removal of the base by DNA-glycosylases results in apurinic or apyrimidinic (AP) sites. The AP-sites are further processed by AP-endonucleases, which excise the remaining sugar backbone. Subsequently, an undamaged nucleotide is inserted into the resulting gap by DNA polymerase I and sealed through the action of a ligase (David *et al.*, 2007; Huffman *et al.*, 2005).

MMR removes incorrect bases incorporated during DNA replication (Iyer *et al.*, 2006). In prokaryotes, this repair mechanism is based on the activity of the MMR proteins MutS, MutL, MutH and of the helicase UvrD. The repair process is initiated by binding of MutS to mismatched base pairs with subsequent recruitment of MutL. This complex activates MutH, which binds to the DNA and cleaves the strand 5' to a GATC site. Subsequently, the segment from the cleavage site to the mismatch is removed by UvrD. The gap is filled by DNA polymerase I and sealed by DNA ligase (Kunkel and Erie, 2005). Correction of the replication error requires that the MMR enzymes specifically recognise the newly synthesised strand. In *E. coli*, this is achieved by methylation of all adenines that occur within GATC sequences (Schlagman *et al.*, 1986). Immediately after DNA replication, the newly synthesised strand is not methylated yet and could thus be specifically recognised by the MMR machinery. MMR proteins were first identified by genetic analyses of *Escherichia coli* mutator mutants. Inactivation of *mutS* or *mutL* genes increases spontaneous mutation by a factor of 50-100, and is the most common cause of an increased spontaneous mutation rate in *E. coli* (LeClerc *et al.*, 1996).

Repair of DNA double-strand breaks

In addition to the various modes of base damage that can arise, cells may experience breakage of one or both chains of the DNA duplex (Friedberg, 2005). The main mechanism for repairing double-strand breaks is homologous recombination repair. This repair pathway takes advantage of the fact that each molecule has an identical copy present in the cell following replication. Damage to one DNA molecule can therefore be repaired using its identical copy as a template without loss of any genetic information (Ishino *et al.*, 2006; Wyman *et al.*, 2004). An alternative mechanism, called nonhomologous end-joining (NHEJ) joins broken chromosome ends independently of homologous sequences at the price of error incorporation (Burma *et al.*, 2006).

SOS response

Most bacteria respond to DNA damage by inducing a multitude of genes involved in repair and control of cell division. This so-called SOS-response is regulated by the global repressor protein LexA, and depends on the action of a functional *recA* gene product (Cox *et al.*, 2000). In a *recA* mutant strain, genes regulated by LexA are no longer DNA damage inducible (Schlacher *et al.*, 2006).

DNA damage tolerance

Besides the DNA damage repair strategies detailed above, cells have evolved a mechanism to tolerate damage. Among the genes that are induced during the SOS response are low-fidelity or "mutator" polymerases, which are able to bypass DNA lesions. However, the benefit of overcoming the replication block is contrasted by the introduction of mutations in the newly synthesised sequence (Goodman, 2002; Nohmi, 2006).

1.2.2 DNA repair in mycobacteria

In silico genome analyses of mycobacterial genomes, i.e. *M. tuberculosis* (Cole *et al.*, 1998; Mizrahi and Andersen, 1998), *M. leprae* (Cole *et al.*, 2001), *M. bovis* (Garnier *et al.*, 2003), *M. avium, M. paratuberculosis* and *M. smegmatis* (The Institute for Genome Research; http://www.tigr.org) indicate the presence of genes coding for enzymes involved in damage reversal, NER, BER, recombinational repair, NHEJ and SOS repair. However, mycobacteria are devoid of the otherwise highly conserved MMR proteins MutS and MutL (Mizrahi and Andersen, 1998).

Biological evidence for the lack of a classical mismatch repair function in mycobacteria was provided by testing mononucleotide repeat instability in *M. smegmatis*. *M. smegmatis* generated frameshifts with a much higher rate than general mutations, indicating the absence of a replication-associated mononucleotide repeat stabilizing function. However, despite the decisive role of MMR in ensuring accuracy of DNA replication and recombination, mycobacteria do not show the high mutation rates typical for mutants lacking MMR (Springer *et al.*, 2004). In addition, *M. tuberculosis* is able to resist extensive periods of exposure to ROI and RNI during persistent infection, which amongst other targets damage DNA (Zahrt and Deretic, 2002). Taken together, these observations indicate that mycobacteria possess efficient mechanisms to maintain genome integrity. Most likely, other DNA repair pathways have to compensate for MMR deficiency.

The most frequent types of DNA damage resulting from exposure to RNI and ROI are thought to be alterations in DNA bases and deoxyribose damages (Demple and Harrison, 1994), which are subject to repair by excision repair pathways like BER and NER. However, the role of excision repair pathways in the pathogenesis and persistence of mycobacteria is not yet fully understood. Little progress has been made so far in understanding the role of BER in

mycobacteria. Using a strategy based on high-throughput screening of a transposon library, several BER enzymes have been identified to be essential for growth *in vivo*, implying a role in virulence (Sassetti *et al.*, 2003; Sassetti and Rubin, 2003). Furthermore, *in silico* analyses suggest that mycobacteria possess at least ten different DNA glycosylases, although the *in vivo* role of most of these homologues in BER remains uncertain (Mizrahi and Andersen, 1998). Research on the uracil DNA glycosylase (Ung) of *M. smegmatis* suggests an important role in mutation prevention and tolerance to nitrosative stress. Survival of the *M. smegmatis ung* mutant in a macrophage infection model was significantly impaired compared to the wildtype, alluding to a role of BER in mycobacterial pathogenicity (Venkatesh *et al.*, 2003). Involvement of NER in resistance of *M. tuberculosis* against nitrosative and oxidative stress has been reasoned by a study that employed a transposon mutagenesis strategy as well. In a screen for mutants that became hypersusceptible to acidified nitrite, transposon mutations in the *uvrB* gene of *M. tuberculosis* have been identified. Further analyses in a mouse model of infection revealed a reduced capacity of these mutants to resist ROI and RNI *in vivo*, indicating a role of NER in *M. tuberculosis* persistence (Darwin and Nathan, 2005).

Intracellular *M. tuberculosis* is exposed to a variety of DNA damaging agents and experiences an accumulation of mutations and double-strand breaks, in particular during the latent phase. Repair of double-strand breaks is an essential process in maintaining genome integrity and is a prerequisite for resuscitation from dormancy (Stephanou *et al.*, 2007). DNA double-strand breaks can be repaired either by homologous recombination or by NHEJ. The homologous recombination repair pathway is not required for survival of *M. tuberculosis* in a mouse model of acute infection (Sander *et al.*, 2001a), which led to the assumption that NHEJ might provide a possible alternative for repair of double-strand breaks during persistence. Proteins involved in that pathway have been characterised recently in *M. tuberculosis*. These *in vitro* analyses demonstrated a functional role of mycobacterial Ku, Ligase D and Ligase C in the error-prone repair of double-strand breaks (Gong *et al.*, 2005; Pitcher *et al.*, 2007). Interestingly, recent investigations revealed that NHEJ of *M. smegmatis* is specifically required for the repair of DNA double-strand breaks in late stationary phase (Stephanou *et al.*, 2007). However, the *in vivo* role of the NHEJ repair pathway in pathogenesis and persistence of *M. tuberculosis* remains yet to be elucidated.

Taken together, these findings indicate that mycobacteria are endowed with a variety of repair pathways that might functionally substitute for each other. The possible redundancy of repair

mechanisms is corroborated by studies on the mycobacterial SOS repair system. Analysis of the regulation of *recA* gene expression revealed that two mechanisms exist for the induction of gene expression following DNA damage in *M. tuberculosis*. Whereas one of these depends on RecA and LexA in the known manner of the classical SOS response, the other mechanism is independent of both of these proteins and induction occurs in the absence of RecA (Davis *et al.*, 2002). Interestingly, most of the genes with known or predicted functions in DNA repair do not depend on RecA function for induction (Rand *et al.*, 2003). These observations point towards the existence of an alternative gene regulation mechanism in response to DNA damage in *M. tuberculosis*.

In summary, the mycobacterial DNA repair system differs in several aspects from that of described model species. Further investigations on the mechanisms that maintain genome integrity in mycobacteria might provide fundamental new insights, not the least with respect to dormancy and the development of drugs for the treatment of the latent state.

1.2.3 Nucleoid structure and maintenance of genome integrity

A common principle to provide and maintain chromosome structure comprises condensation of the DNA into a structure called the nucleoid. Diverse DNA binding proteins are associated with the nucleoid and take part in chromosome organisation to assist in diverse and complex processes such as replication, recombination, repair, modification and DNA transcription.

Among these proteins are the SMC (structural maintenance of chromosomes) proteins that are conserved from prokaryotes to eukaryotes (Cobbe and Heck, 2000; Cobbe and Heck, 2004; Soppa, 2001). The SMC family proteins are large proteins in the range between 110-170 kDa and share common principles in domain organisation: globular N- and C-terminal domains which are connected by two long coiled coil domains, separated by a globular hinge domain of approximately 150 amino acids in length (Hirano, 2005).

Eukaryotes possess at least six distinct SMCs that function in chromosome condensation, sister chromatid cohesion, DNA repair and dosage compensation (Jessberger, 2002; Losada and Hirano, 2005). Genetic and biochemical evidence suggests that eukaryotic SMCs are also involved in the repair of double-strand breaks (Lehmann, 2005).

In contrast to the preserved multiple SMC proteins in eukaryotes, most bacteria possess only a single SMC (Soppa, 2001). SMC proteins from *Bacillus subtilis* or *Caulobacter crescentus* appear to have functions that resemble those of eukaryotic SMCs. The proteins are not essential but null mutations of SMC result in anucleate cell ("titan cells") and aberrant nucleoid formation, i.e. decondensed chromosomes (Britton *et al.*, 1998; Jensen and Shapiro, 1999; Volkov *et al.*, 2003). Furthermore, deletion of SMC leads to increased susceptibility to DNA damaging agents, indicating a role of prokaryotic SMCs in DNA repair (Dervyn *et al.*, 2004).

Little is known about mycobacterial SMCs. *In silico* analyses revealed that all mycobacterial genomes that have been sequenced so far harbour a single SMC homologue. It is even present in *M. leprae*, a bacterium that has lost various gene functions during reductive genomic evolution (Cole *et al.*, 2001).

2 Aim of this study

2.1 Role of NER in mycobacteria

As part of its intracellular lifestyle, *M. tuberculosis* encounters various adverse conditions, which demand efficient mechanisms to maintain genome integrity. Mycobacteria are devoid of the highly conserved MMR system, which is fundamental to prevent mutations during replication. The finding that mycobacteria exhibit a general mutation rate that is comparable to that of MMR proficient species suggests that mycobacteria use alternative DNA repair systems to compensate for the lack of MMR. Towards this aim, we investigated the genetics and function of the mycobacterial NER pathway. NER knockout mutants were constructed in *M. smegmatis* using allelic replacement techniques. *M. smegmatis*, like *M. tuberculosis*, lacks the MutSLH based mismatch repair system (www.tigr.org). *M. smegmatis* has a full set of homologues for the UvrABCD (NER) pathway; strains containing deletions in the excinuclease component *uvrB* and in the helicase component *uvrD* were examined. To evaluate the significance of the results obtained in *M. smegmatis*, i.e. to assess whether NER is also important in the persistence of the relevant pathogen *in vivo*, strains deficient in the NER excinuclease part *uvrA* and in the NER helicase *uvrD* were constructed in *M. tuberculosis* and studied in *in vivo* infection experiments.

2.2 Role of DNA repair in *M. tuberculosis* pathogenesis

To gain insight into the role of DNA repair mechanisms in the pathogenesis of *M. tuberculosis* infection, we set out to generate *M. tuberculosis* strains with multiple deficiencies in DNA repair pathways. We intended to construct these mutants successively, starting with single mutants, followed by double and triple mutants.

In particular, we aimed to construct strains lacking NER and recombinational repair. To inactivate the recombinational repair pathway, the gene encoding for RecA was targeted for deletion in NER mutant strains using allelic replacement techniques.

A second gene with significant homology to *uvrD*, annotated as *uvrD2*, is present in the *M. tuberculosis* genome. To assess the degree of functional redundancy of these two genes, we set out to construct mutants lacking *uvrD* and its homologue *uvrD2*.

2.3 Role of SMC in mycobacteria

Little is known about chromosome stability, organisation, and partitioning in mycobacteria. *Mycobacterium tuberculosis* displays the ability to persist for decades in its host during latent infection. It is an issue of high interest to investigate how the bacterium manages to stabilise its genome during persistence and how it accomplishes proper cell division during reactivation. As detailed in section 1.3.3, SMC plays a major role in maintaining chromosome stability in various organisms and it was also found to be involved in the repair of DNA double-strand breaks. Functional SMC may therefore be crucial for maintenance of genome integrity during mycobacterial dormancy and may have a decisive role during reactivation. To investigate the role of SMC in mycobacteria, we generated and analysed *M. smegmatis* and *M. tuberculosis* mutants deficient in SMC.

3 Materials and Methods

All chemicals were purchased from *Sigma-Aldrich* or *Fluka* if not specially indicated. Molecular methods were performed according to standard protocols (Sambrook, 1989).

3.1 Bacterial strains, growth conditions and conservation

E. coli strain XL1 blue (*Stratagene*) was used for cloning and propagation of plasmids. Bacterial cultures were grown in Luria Bertani (LB) medium at 37°C for 14-20 h.

Cultures of *M. smegmatis* SMR5 (Sander *et al.*, 1995) and *M. smegmatis* SMR5 *rrnB* (Sander *et al.*, 1996), both derivatives of *M. smegmatis* mc^2 155, were grown either in LB-Tween 80 or in 7H9 medium at 37°C for 2-3 days. *M. smegmatis* SMR5 *rrnB* harbours only a single functional rRNA operon which is a prerequisite for the assays applied in this study.

M. tuberculosis H37Rv #1424, a derivative of *M. tuberculosis* H37Rv (Raynaud *et al.*, 2002), was grown in OADC enriched Middlebrook 7H9 medium at 37°C for 3 weeks in settling cultures.

The above-mentioned parental strains of *M. tuberculosis* and *M. smegmatis* carry a *rpsL* mutation conferring streptomycin resistance.

When appropriate, antibiotics were added at the following concentrations:

Ampicillin	100 µg/ml medium
Hygromycin	100 µg/ml medium
Kanamycin	50 µg/ml medium
Gentamicin	5 µg/ml medium
Streptomycin	100 µg/ml medium
Apramycin	20 µg/ml medium
Clarithromycin	50 µg/ml medium

E. coli were frozen and stored at –80°C in LB medium containing 10% glycerol.
Mycobacteria were stored at –80°C using freezing medium for mycobacteria.

3.2 Growth media

Luria-Bertani Broth (LB)	1% (w/v) Bacto™ Tryptone, *Difco* 0.5% (w/v) Bacto™ Yeast Extract, *Difco* 1% (w/v) sodium chloride
LB-Tween	LB + 0.05% (v/v)Tween
LB-Agar	LB + 1.5% (w/v) Bacto-agar, *BBL*
Middlebrook 7H9	0.47 % (w/v) Middlebrook 7H9 broth, *Difco* 0.2% (v/v) glycerol, *Merck* 0.05% (v/v) Tween
Middlebrook 7H9 OADC enriched	0.47 % (w/v) Middlebrook 7H9 broth, *Difco* 0.2% (v/v) glycerol, *Merck* 0.05% (v/v) Tween Middlebrook OADC (oleic acid albumin dextrose) enrichment, *Difco*
Middlebrook 7H10-Agar	Middlebrook 7H10 agar, *Difco* 0.4% (v/v) glycerol, *Merck*
Middlebrook 7H10-Agar OADC enriched	Middlebrook 7H10 agar, *Difco* 0.4% (v/v) glycerol, *Merck* Middlebrook OADC (oleic acid albumin dextrose) enrichment, *Difco*
Mueller-Hinton agar / 5% sheep blood	2.0 g/l meat infusion 17.5 g/l casein hydrolysate, 1.5 g/l starch, 65 ml sheep blood, 14.0 g/l agar-agar
Freezing medium for mycobacteria	2.6 g Middlebrook 7H9 broth, (*Difco*), 113 g saccharose, 2.22 ml glycerol, dist. water (1000 ml)

3.3 Buffers and solutions

3.3.1 General

EB-buffer	10 mM Tris/HCl, pH 8.5
DNA loading buffer (6x)	15 % (w/v) Ficoll Type 400
	0.25 % (w/v) Bromphenolblue/ Xylenecyanol
PBS (phosphate buffer saline)	220 mM NaCl, 150 mM Na_2HPO_4
	130 mM KH_2PO_4, pH 7.4
Physiological NaCl-solution	0.9 g NaCl, dist. water (1000 ml)
SDS	10 % SDS (w/v)
TAE-buffer (50 x)	2 M Tris base, pH 8, 0.1 M EDTA
TE-buffer	10 mM Tris/HCl, pH 7.5, 1 mM EDTA

3.3.2 Southern blot

HCl solution	0.25 M HCl
Denaturation solution	1.5 M NaCl, 0.5 M NaOH
Neutralization solution	0.5 M Tris Base, 3 M NaCl, pH 7.5
20 x SSC	3 M NaCl, 0.3 M sodium citrate, pH 7.0
Wash buffer	0.1 M maleic acid, 0.15 M NaCl, 0.3% (v/v) Tween20
Maleic acid buffer	0.1 M maleic acid, 0.15 M NaCl, pH 7.5
Detection buffer	0.1 M NaCl, 2 M Tris-base, pH 9.5
Hybridization solution	0.05 M SDS, 50% formamide (v/v), 25% 20 x SSC (v/v),
	5% 1 M $NaPO_4$ pH 7.0 (v/v), 0.003 M N-lauroylsarcosine,
	20% (v/v) 10x blocking solution
Blocking solution	10% blocking solution (10 x), 90% maleic acid buffer
Stripping buffer	0.2 M NaOH, 0.1% (v/v) SDS

3.3.3 Assays: Stock solutions

Rifampicin: 20 mg/ml in dimethylsulfoxide (DMSO)

Mitomycin C: 0.2 mg/ml in water

tert-Butyl-hydroperoxide: 70 % solution in water

Sodium nitrite: 3 M solution in water

3.4 Primers and plasmids used in this study

Table 1: Primers (f = forward primers, r = reverse primers)

	Restriction site	Gene specific	Tm	Function
1f	*GGAATTC*CATATG *NdeI*	GTGAGGACGCCTACGAC	56°C	Amplification of upstream region of *M. smegmatis uvrD1*
1r	*GGA*AGATCT *BglII*	GGTGAACGTGATGGCCAG	58°C	
2f	*GGA*AGATCT *BglII*	GCAGGAACTCATCGACTGGCG	68°C	Amplification of downstream region of *M. smegmatis uvrD1*
2r	*TGC*ATGCAT *NsiI*	GATCGCGTCGGGCACCTTC	68°C	
3f	*GGAATTC*CATATG *NdeI*	GGCCGAGTACGGCCAGTC	58°C	Amplification of upstream region of *M. smegmatis uvrB*
3r	*GGA*AGATCT *BglII*	GCGCCATCACGAGCGTG	58°C	
4f	*GGA*AGATCT *BglII*	GAGTCGGTCGAGATCGGTGG	66°C	Amplification of downstream region of *M. smegmatis uvrB*
4r	*TGC*ATGCAT *NsiI*	CGTCATGTGCGCCAGCCGC	66°C	
5f	*GGAATTCCATATG* *NdeI*	CGACACCGTCCTGCCG	56°C	Amplification of upstream region of *M. smegmatis smc*
5r	*GGAAGATCT* *BglII*	CGTCTCGGCCTGATTGG	56°C	
6f	*GGAAGATCT* *BglII*	CACCGCGCTCAACAACC	56°C	Amplification of downstream region of *M. smegmatis smc*
6r	*TGCATGCAT* *NsiI*	GGCACCTTCTGTCACCG	56°C	
7f	*GGAATTC*CATATG *NdeI*	GAACTGGCATGACAACGCACG	66°C	Amplification of upstream region of *M. tuberculosis uvrD2*
7r	*GGA*AGATCT *BglII*	CACGGCGAACTTGCTGTCCAG	66°C	
8f	*GGA*AGATCT *BglII*	GACGTGTGCAGCCGACG	58°C	Amplification of downstream region of *M. tuberculosis uvrD2*
8r	*TGCATGCAT* *NsiI*	CCAGGTCGAGGTCTTCC	56°C	
9f	*GGAATTC*CATATG *NdeI*	CCCTCAAGTCGGGCACCAC	64°C	Amplification of upstream region of *M. tuberculosis smc*
9r	CTAGCTAGC *NheI*	CCATCGTGTCGAGTTTGCGCA	66°C	
10f	GCTAGC *NheI*	GACGAGGTGAACACGCTGAG	64°C	Amplification of downstream region of *M. tuberculosis smc*
10r	CCATGG *NcoI*	CAGCGCAGCGATGACGACC	64°C	

Table 2: Sequencing Primers

T7	5′-TAA TAC GAC TCA CTA TAG GG-3′
SP6	5′-ATT TAG GTG ACA CTA TAG-3′
pMCS5 MCS *Sma*I → *Nde*I	5′-TAC GTA GGT ACC GAG CTC-3′
5' pMCS5-122	5′-GTT TTC CCA GTC ACG ACG-3′
***Hyg* # 692R:**	5′-GTT CTCGGT GGT G-3′
***Hyg* # 834F:**	5′-CAC GGG ACC AAC ATC TTC G-3′
5' *rpsL* out	5′-TGT CGT CCA TGA CGC CGG TC-3′
3' *rpsL* out	5′-GCA GGG TGT CAA GAA CCG-3′

Table 3: Plasmids

Parental vectors for cloning	
pGEM®-T Easy	(*Promega*)
pBluescript II KS (-)	(*Stratagene*)
pBluescript-*rpsL*	(Sander *et al.*, 2001a)
pUC4K	(*Pharmacia*)
pML10	(Labes *et al.*, 1990)
ptrpA1-*rpsL*	(Sander *et al.*, 2001b)
pMCS5	(*Stratagene*)
pMCS5-*rpsL*	(Rezwan, 2006)
pMCS5-*rpsL-hyg*	(this study)
pMV 361	(Stover *et al.*, 1991)

Suicide vectors for generation of knockout mutants		
p*uvrD-rpsL*-2-II	Disruption of *M. tuberculosis uvrD*	(this study)
p*uvrA*::*aph-rpsL*	Disruption of *M. tuberculosis uvrA*	(this study)
p*recA-hyg*	Disruption of *M. tuberculosis recA*	(Sander *et al.*, 2001a)
pMCS5-*rpsL-hyg-uvrD2*	Disruption of *M. tuberculosis uvrD2*	(this study)
pMCS5-*rpsL-smc-hyg*	Disruption of *M. tuberculosis smc*	(this study)
pMCS5-*rpsL-hyg-uvrD* (sm)	Disruption of *M. smegmatis uvrD*	(this study)
pMCS5-*rpsL-hyg-uvrB* (sm)	Disruption of *M. smegmatis uvrB*	(this study)
pMCS5-*rpsL- smc* (sm)-*hyg*	Disruption of *M. smegmatis smc*	(this study)

Integrative vectors for complementation of knockout mutants		
pMV361-*uvrD*	Complementation of *M. tuberculosis uvrD*	(this study)

Recombination substrates containing a *rrnA* fragment All vectors contain a mutated *rrnA* fragment which confers antibiotic resistance (denoted in parentheses) upon recombination with the chromosomal wildtype *rrnA* gene. Numbers indicate the *rrnA* position of the mutated nucleotide.	
pMV361ΔKan-Gm-rRNA A 2059 C (Clarithromycin)	(Pfister *et al.*, 2004)
pMV361ΔKan-Gm-rRNA A 2059 G (Clarithromycin)	(Pfister *et al.*, 2003)
pMV361ΔKan-Gm-rRNA A 2058 C (Clarithromycin)	(Pfister *et al.*, 2004)
pMV361ΔKan-Gm-rRNA A 2058 G (Clarithromycin)	(Pfister *et al.*, 2004)
pMV361ΔKan-Gm-rRNA G 1491 A (Apramycin)	(Pfister *et al.*, 2005)
pMV361ΔKan-Gm-rRNA G 1491 C (Apramycin)	(Pfister *et al.*, 2005)
pMV361ΔKan-Gm-rRNA G 1491 T (Apramycin)	(Pfister *et al.*, 2005)

3.5 Preparation of competent bacteria and transformation

3.5.1 *Escherichia coli*

An overnight culture of *E. coli* strain XL-1 blue was inoculated into one litre of LB medium and cultivated at 37°C until OD_{600} reached 0.5-0.6. To prepare electrocompetent bacteria, cells were collected by centrifugation at 5000 x g at 4°C for 15 min and washed twice with an equal volume ice-chilled 1 mM HEPES (pH 7.0). The washing procedure was repeated twice with an equal volume of 10% glycerol. Finally, bacteria were resuspended in 5-10 ml of 10% glycerol. The suspension was divided into 40 µl aliquots, frozen in liquid nitrogen and stored at –80°C until use.

For electroporation, 40 µl of competent bacteria were mixed with 2 µl of precipitated ligation reaction or 0.1 µg of supercoiled plasmid DNA (transformation control), respectively, and kept on ice for 5 min. Electroporation was performed in 0.1 cm cuvettes with a single pulse [1.8 kV, 2 µF, 200 Ω (Gene Pulser Xcell™, *BioRad*)]. Bacteria were immediately resuspended in 1 ml of LB medium and incubated at 37°C for 1 h. Transformants were selected by plating 100 µl of electroporated bacteria on LB agar plates containing an appropriate antibiotic. Plates were incubated at 37°C for 14-20 h.

3.5.2 *Mycobacterium smegmatis*

M. smegmatis strains were cultivated in LB-Tween medium. When an OD_{600} of 0.6-0.8 was achieved, bacteria were incubated on ice for 30 minutes. Subsequently, the bacteria were collected by centrifugation at 5000 x g at 4°C for 15 min, washed several times with ice-chilled 10% glycerol and finally resuspended in 5 ml glycerol. The suspension was divided into 410 µl aliquots, frozen in liquid nitrogen and stored at -80°C until use.

For electroporation, competent bacteria were thawed on ice and washed once with ice-chilled 10% glycerol. Subsequently, 100 µl competent bacteria were mixed with 1 µg of plasmid DNA and placed on ice for 10 min. Electroporation was performed in 0.4 cm cuvettes with a single pulse [2.5 kV, 25 µF, 1000 Ω (Gene Pulser Xcell^{TM}, *BioRad*)]. Bacteria were immediately resuspended in 1 ml of LB-Tween 80 medium and incubated at 37°C for 3 h. Electroporated bacteria were selected by plating 100 µl on LB agar plates containing an appropriate antibiotic. Plates were incubated at 37°C for 3-4 days.

3.5.3 *Mycobacterium tuberculosis*

M. tuberculosis strains were cultivated in Middlebrook 7H9 medium enriched with OADC. When an OD_{600} of 0.7-1.0 was achieved, glycine (2 M) was added (final concentration 1.5%) to increase transformation efficiency. After incubation for additional 20-24 h, the bacteria were harvested by centrifugation, washed several times with 10% glycerol, and bacteria were finally resuspended in a volume of 5 ml glycerol. For electroporation, 400 µl of competent cells were mixed with 1 µg of supercoiled plasmid DNA and electroporated (Gene pulser II, *Bio-Rad*) with following settings: 2.5 kV, 1000 Ω, 25 µF. After electroporation, cells were resuspended in 4 ml of 7H9-OADC and incubated for 24 h at 37°C. Following incubation, appropriate dilutions were plated on selective agar. Plates were incubated at 37°C for 3-6 weeks.

3.6 DNA manipulations and isolation of plasmids

Standard techniques were used for DNA manipulation. The restriction endonucleases and DNA modifying enzymes were purchased from *MBI Fermentas*, *Roche* and *Amersham*.
PCR was performed in a DNA thermal cycler (Tpersonal, *Biometra*) with triple master PCR system (*Eppendorf*) or GC rich system (*Roche*) according to the manufacturers' protocols.
All initial cloning procedures were performed with *E. coli* XL1-Blue. Plasmids were prepared with a NucleoSpin®Plasmid preparation kit (*Macherey-Nagel*) or PureYield™ Plasmid Midiprep system (*Promega*) according to the manufacturers' recommendations. Plasmid DNA was dissolved in water in concentrations of 500 to 1,000 ng/µl.

3.7 DNA sequencing

Plasmid DNA was extracted as described in section 3.6. The ABI Dye Terminator Cycle Sequencing Kit was used for sequencing on an ABI Prism 310 Genetic Analyzer (*Applied Biosystems*) following the supplier's instruction. Sequence analyses were performed using sequence analysis software ABI Prism 310 collection, Sequencing Analysis Version 3.4.1.

3.8 Isolation of genomic DNA

Bacteria were grown for 3-5 days (*M. smegmatis*) or 3-4 weeks (*M. tuberculosis*). Bacteria were resuspended in 340 µl TE buffer and heat inactivated at 80°C for 20 min. After cooling down to room temperature, 2 µl 20 % Tween 80 and 10 µl lysozyme (80 mg/ml, *Roche*) were added, followed by incubation at 37°C / 300 rpm for 2 hours. After addition of 20 µl 20 % SDS and 20 µl proteinase K (2 mg/ml, *Roche*), samples were incubated at 50°C for 1 hour. Phenol / chloroform / isoamylalcohol extraction was performed by adding 400 µl of a (25/24/1) mixture with subsequent centrifugation at 16,000 g at 4°C for 20 min and transfer of the supernatant into a 1.5 ml tube. 8 µl 5 M NaCl and 2.5 volumes (1 ml) of ethanol were added and the solution was incubated at -20°C overnight. After centrifugation of the samples at 13,000 rpm at 4°C for 20 min, the pellet was washed twice with 70 % ethanol, dried under vacuum and resuspended in 100 µl EB buffer.

3.9 Southern blot

For Southern blot analyses, 200 ng of genomic DNA was digested with appropriate restriction enzymes, separated by agarose gel electrophoresis, and treated according to standard protocols. DNA was transferred to a positively charged nylon membrane (*Roche*) by vacuum blotting using a VacuBlot System (*Biometra*). Subsequently, the DNA was cross-linked to the membrane by UV- irradiation (UV–Stratalinker, *Stratagene*). DNA was hybridised to a digoxigenin (DIG)-labelled probe, washed under stringent conditions, and detected with an anti-digoxigenin antibody coupled with horseradish peroxidase (*Roche*).

Labelling of probes

Templates were excised from the corresponding vectors with appropriate restriction nucleases that create fragments with a size of 200-1000 bp. Nucleic acid probes were generated by labelling with digoxigenin (DIG) according to the manufacturer's recommendation.

3.10 Gene disruption in *M. tuberculosis* and *M. smegmatis*

To generate *M. tuberculosis* and *M. smegmatis* knockout mutants, allelic replacement techniques (deletion mutagenesis) were applied using a combination of a suicidal vector and counterselection with streptomycin (Sander et al 1995). Resistance to streptomycin in mycobacteria involves mutation within the *rpsL* gene coding for the ribosomal protein S12 (Finken *et al.*, 1993). Sensitivity to streptomycin is dominant over resistance in a merodiploid strain carrying a wildtype allele and a mutated allele (Kenney and Churchward, 1994; Lederberg, 1951). A suicidal vector in which a wildtype *rpsL* gene is cloned in proximity to a target gene (mutagenized by deletion of parts of the open reading frame) was used. In case of unmarked deletion mutagenesis, the deletion allele was flanked by a *hyg* cassette coding for resistance to hygromycin. In case of marked deletion mutagenesis, the respective resistance cassette was cloned into the deletion allele. The gene replacement vector was transformed into a streptomycin-resistant derivative of *M. smegmatis* or *M. tuberculosis*, with a mutated *rpsL*. The first selection for maintenance of cloned sequences was conferred by the positive selectable marker (e.g. the resistance cassette). Transformants were purified by streaking on selective agar containing the respective antibiotic. Individual clones were investigated by PCR and Southern blot analysis for integration by homologous recombination (single cross-over). Single cross-over transformants were grown up in liquid broth and subsequently

subjected to a second selection step. For counterselection streptomycin was used to select against maintenance of plasmid sequences which contain the gene encoding the streptomycin-sensitive allele of the *rpsL* gene (Figure 1). Putative knockout mutants which have undergone the counterselection procedure were colony purified and investigated by PCR and Southern blot analysis for allelic replacement recombinants.

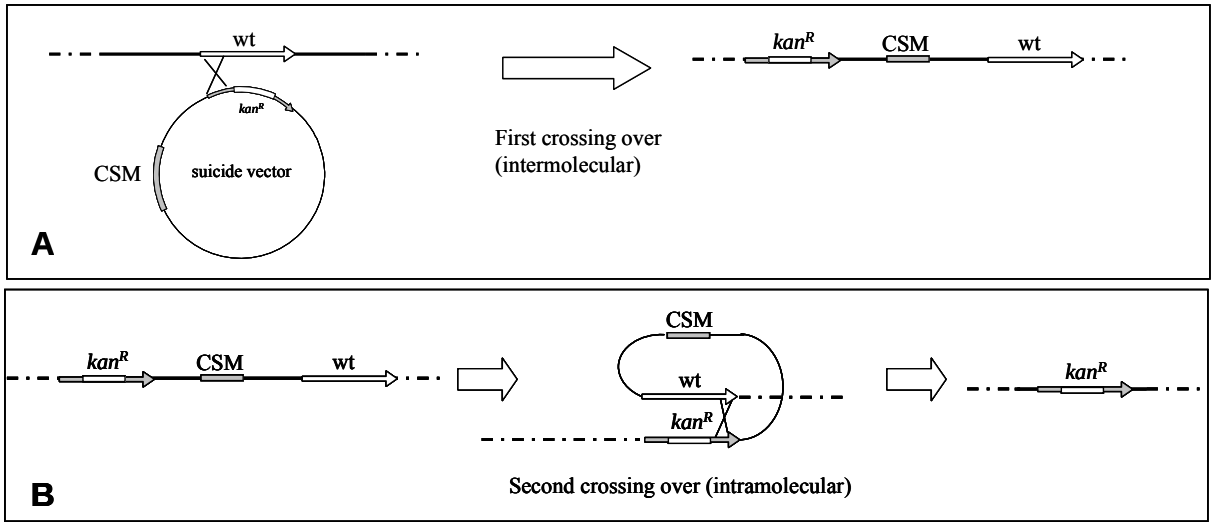


Figure 1: Targeted allelic replacement strategy

Shown is the strategy for marked deletion mutagenesis making use of a kanamycin resistance cassette **(A)** First crossing over. The suicide vector carries the counterselectable marker (CSM) *rpsL* and the mutated gene interrupted by a positive selectable marker (*kan*^{R}); wt = wildtype target gene of the recipient. **(B)** Second crossing over leading to inactivation of the target gene.

Cloning strategy for generation of knockout vectors

Using genomic DNA, 1-1.5 kbp fragments upstream (5') and downstream (3') of the predicted open reading frame were amplified. The primers were constructed to contain restriction sites (Table 1). The PCR fragments were subcloned into pGEM®-T Easy Vector and the accuracy of the PCR product was controlled by sequencing with primers T7 and SP6. Subsequently, the PCR fragments were cloned into the suicide vector pMCS5-*rpsL-hyg* for unmarked knockouts (representative scheme see Figure 2), or into pMCS5-*rpsL* for marked knockouts.

The resulting knockout vectors were controlled by sequencing. The mentioned cloning protocol was applied to the construction of all knockout vectors if not otherwise indicated.

Generation of *M. smegmatis* mutants

Disruption of *M smegmatis uvrD* (unmarked)

uvrD open reading frame: 562 5153 – 562 2802 (2352 bp)

Cloned 5' flanking region: 562 6132 – 562 4935 (1198 bp), amplified with primers 1f and 1r

Cloned 3' flanking region: 562 3114 – 562 1904 (1210 bp), amplified with primers 2f and 2r

Deleted region: 562 4934 – 562 3115 (1821 bp)

The suicide vector pMCS5-*rpsL-hyg- uvrD* (sm) was constructed according to the scheme shown in Figure 2 a.

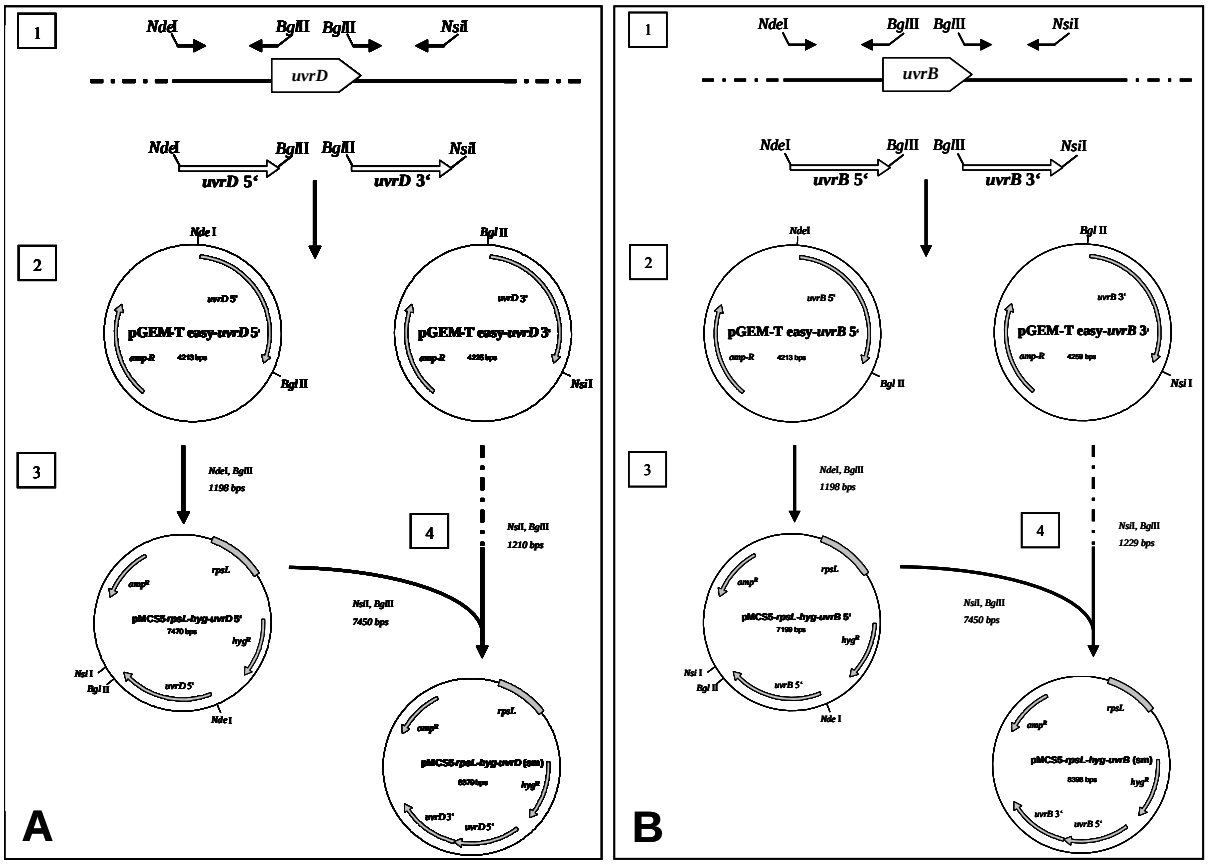


Figure 2: Cloning of knockout vectors for the generation of unmarked mutants

(A) Knockout vector for *M. smegmatis uvrD* **(B)** Knockout vector for *M. smegmatis uvrB*

Using genomic DNA, 1-1.5 kbp fragments upstream (5') and downstream (3') of the predicted open reading frame were amplified. The primers were constructed to contain restriction sites **(1)**. The PCR fragments were subcloned into pGEM®-T Easy vector **(2).** The 5' PCR fragments were cloned into the suicide vector pMCS5-*rpsL-hyg* to result in pMCS5-*rpsL-hyg- uvrD 5'*, and pMCS5-*rpsL-hyg- uvrB 5'*, respectively **(3).** The 3' PCR fragments were added to result in the knockout vectors pMCS5-*rpsL-hyg- uvrD* (sm), and pMCS5-*rpsL-hyg-uvrB* (sm) **(4)**. This strategy was also applied for generation of the knockout vector for *M. tuberculosis uvrD2.*

Disruption of *M. smegmatis uvrB* (unmarked)

uvrB open reading frame: 388 7521 – 388 5361 (2160 bp)

Cloned 5' flanking region: 388 8458 – 388 7260 (1198 bp), amplified with primers 3f and 3r

Cloned 3' flanking region: 388 5632 – 388 4403 (1229 bp), amplified with primers 4f and 4r

Deleted region: 388 7259 – 388 5633 (1628 bp)

The suicide vector pMCS5-*rpsL-hyg- uvrB* (sm) was constructed according to the scheme shown in Figure 2 b.

Generation of *M. smegmatis uvrD / uvrB* double mutant

The *uvrB* knockout vector pMCS5-*rpsL-hyg-uvrB* (sm) was transformed into the *M. smegmatis uvrD* mutant.

Disruption of *M. smegmatis smc*

smc open reading frame: 1 2905 – 1 6492 (3587 bp)

Cloned 5' flanking region: 1 2569 – 1 3652 (1083 bp), amplified with primers 5f and 5r

Cloned 3' flanking region: 1 5565 – 1 6565 (1000 bp), amplified with primers 6f and 6r

Deleted region: 1 3653 – 1 5564 (1911 bp)

The suicide vector pMCS5-*rpsL-hyg- smc* (sm) was constructed according to the scheme shown in Figure 3.

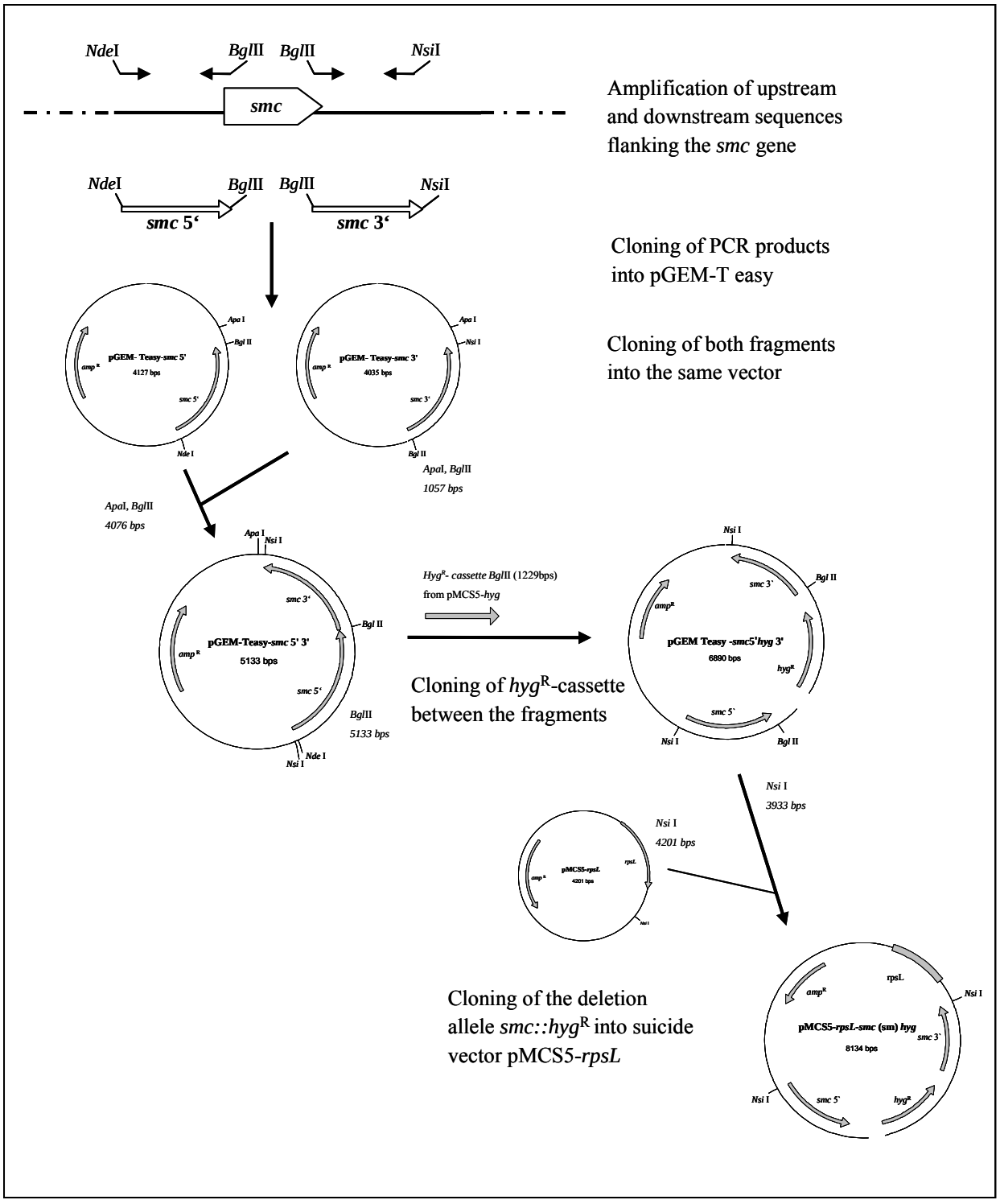


Figure 3: Cloning of the knockout vector for the generation of an *M. smegmatis smc* mutant

Generation of *M. tuberculosis* mutants

Disruption of *M. tuberculosis uvrD*

A 5134 bp *Apa*I/*Eco*RV fragment comprising the *uvrD* gene was isolated from bacterial artificial clone (BAC) *Rv103* (Brosch et al, 1998) and cloned into plasmid ptrpA-1-*rpsL* previously digested with *Apa*I/*Eco*RV resulting in p*uvrD-rpsL*. For functional inactivation of *uvrD* a 1.4 kbp fragment (*Sal*I-*Aat*II) was deleted and substituted by a 3 kbp gentamicin resistance cassette originally derived from plasmid pML10 resulting in plasmid p*uvrD::gm-rpsL* (Figure 4a). This plasmid was transformed into the streptomycin resistant *M. tuberculosis* strain H37Rv #1424.

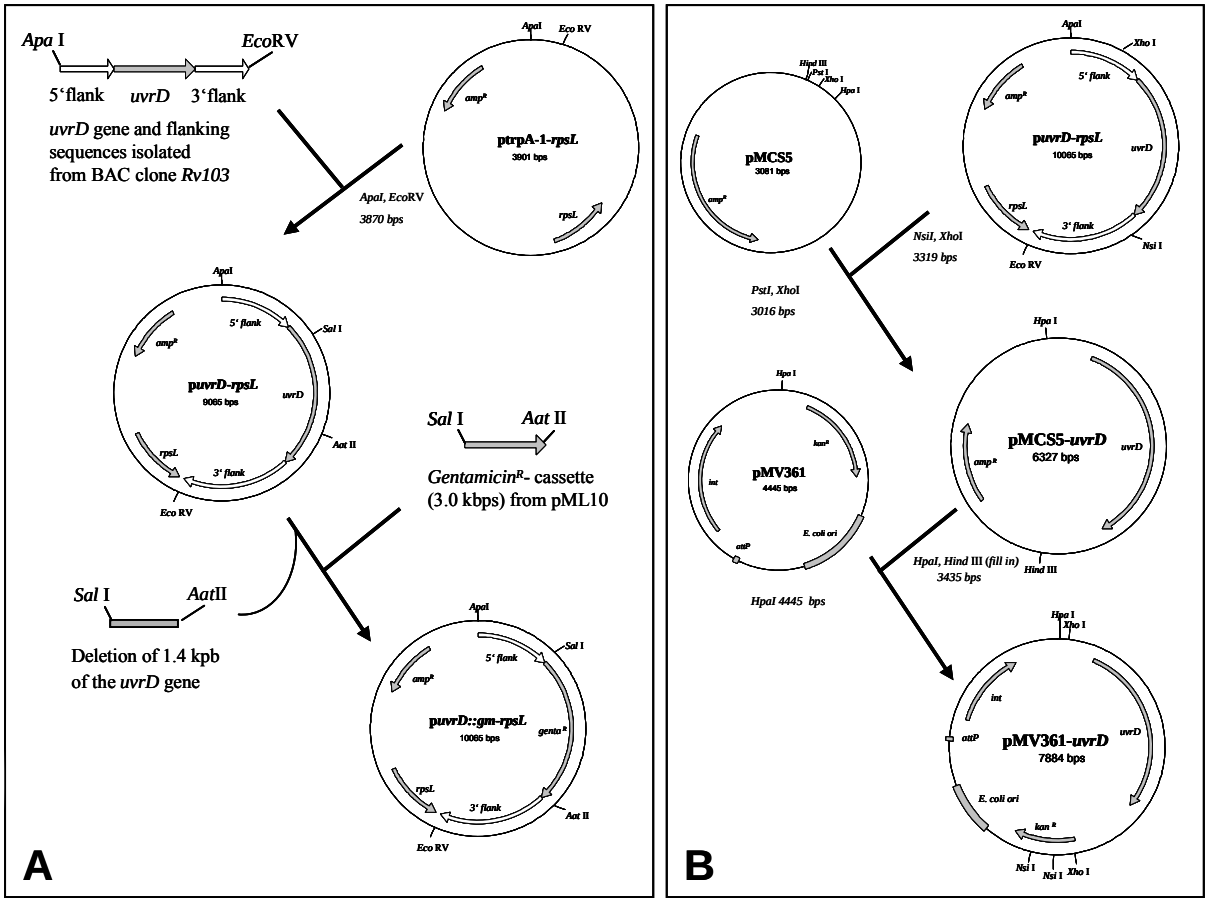


Figure 4: Cloning of vectors for the generation *M. tuberculosis uvrD* mutant (A) and complementation of *M. tuberculosis uvrD* mutant (B)

For complementation vector p*uvrD-rpsL* was digested with *Nsi*I/*Xho*I to obtain a 3319 bp fragment comprising the whole *uvrD* gene plus 362 bp upstream and 641 bp downstream sequences. This fragment was subcloned into vector pMCS5 and subsequently inserted into the *Hpa*I site of integrative vector pMV361 (Figure 4b).

uvrD open reading frame: 105 8260 - 106 0575 (2316 bp)
Cloned region: 105 7112 - 106 2246 (5134 bp)
Deleted region: 105 8559 - 105 9943 (1384 bp)
Complementation: 105 7898 - 106 1216

Disruption of *M. tuberculosis uvrA*

A 5.9 kbp *Not*I fragment comprising the *uvrA* gene was isolated from BAC *Rv401* (Brosch et al, 1998) and cloned into pBluescript KS (-) (*Stratagene*) previously digested with *Not*I resulting in pBluescript-*uvrA*. For functional inactivation of *uvrA* the resulting plasmid was digested with *Eco*NI and *Stu*I (deletion of a 70 bp fragment and of a 1.2 kbp fragment).
Subsequently, a 1.2 kbp fragment containing the kanamycin resistance cassette from plasmid pUC4K was inserted resulting in plasmid p*uvrA::aph*. From this plasmid a 5.9 kbp *Not*I fragment comprising the inactivated *uvrA* gene was subcloned into plasmid pBluescript-*rpsL* (M.s.) previously digested with *Not*I resulting in plasmid p*uvrA::aph-rpsL* (Figure 5). pBluescript-*rpsL* (M.s.) is a derivative of pBluescript carrying a 0.9 kbp fragment with the *rpsL* gene of *M. smegmatis*.

uvrA open reading frame: 184 3741-184 6659 (2919 bp)
Cloned region: 184 1810-184 7701 (5892 bp)
Deleted region: 184 4659-184 6001 (1343 bp)

Generation of *M. tuberculosis uvrA / uvrD* double mutant

The *uvrD* knockout vector was transformed into the *M. tuberculosis uvrA* mutant.

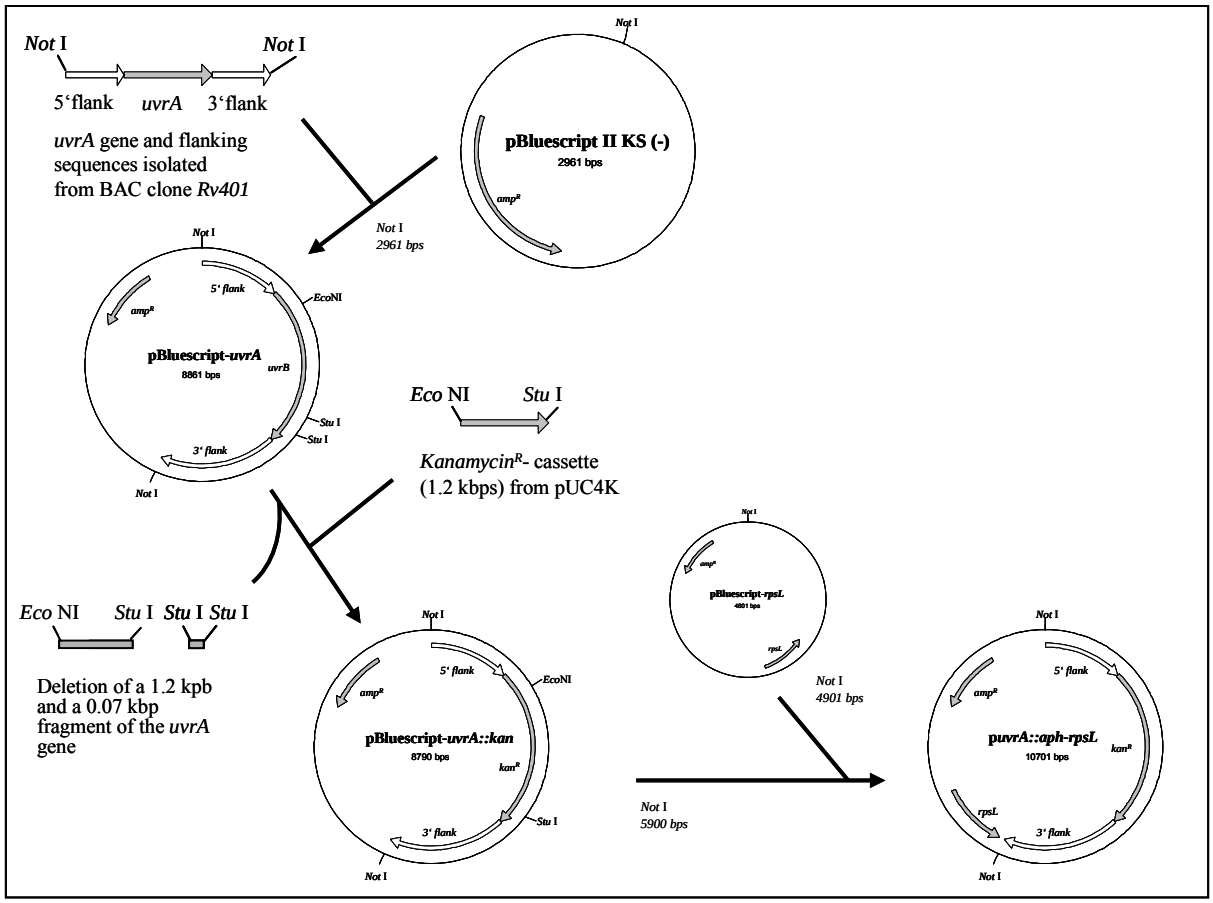


Figure 5: Cloning of the knockout vector for the generation of an *M. tuberculosis uvrA* mutant

Disruption of *M. tuberculosis uvrD2* (unmarked)

uvrD2 open reading frame: 356 9109 - 357 1211 (2103 bp)

Cloned 5' flanking region: 356 8121 - 356 9346 (1225 bp), amplified with primers 7f and 7r

Cloned 3' flanking region: 357 0831 - 357 2148 (1317 bp), amplified with primers 8f and 8r

Deleted region: 356 9347 - 357 0830 (1485 bp)

Construction of the suicide vector pMCS5-*rpsL-hyg-uvrD2* (tb) was similar to the construction shown in Figure 2.

Generation of *M. tuberculosis* mutants with multiple deficiencies in DNA damage repair pathways

To generate multiple knockout mutants, genes were inactivated consecutively. Towards this aim, the *M. tuberculosis uvrD* mutant strain and the double mutant strain *uvrA / uvrD* were transformed with knockout-vectors *recA* and *uvrD2*.

Disruption of *M. tuberculosis smc*

smc open reading frame: 323 4189 - 323 7806 (3618 bp)

Cloned 5' flanking region: 323 4077 - 323 5133 (1056 bp), amplified with primers 9f and 9r

Cloned 3' flanking region: 323 7254 - 323 8138 (884 bp), amplified with primers 10f and 10r

Deleted region: 323 5134 - 323 7253 (2121 bp)

Construction of the suicide vector pMCS5-*rpsL-hyg- smc* (tb): see Figure 6**.**

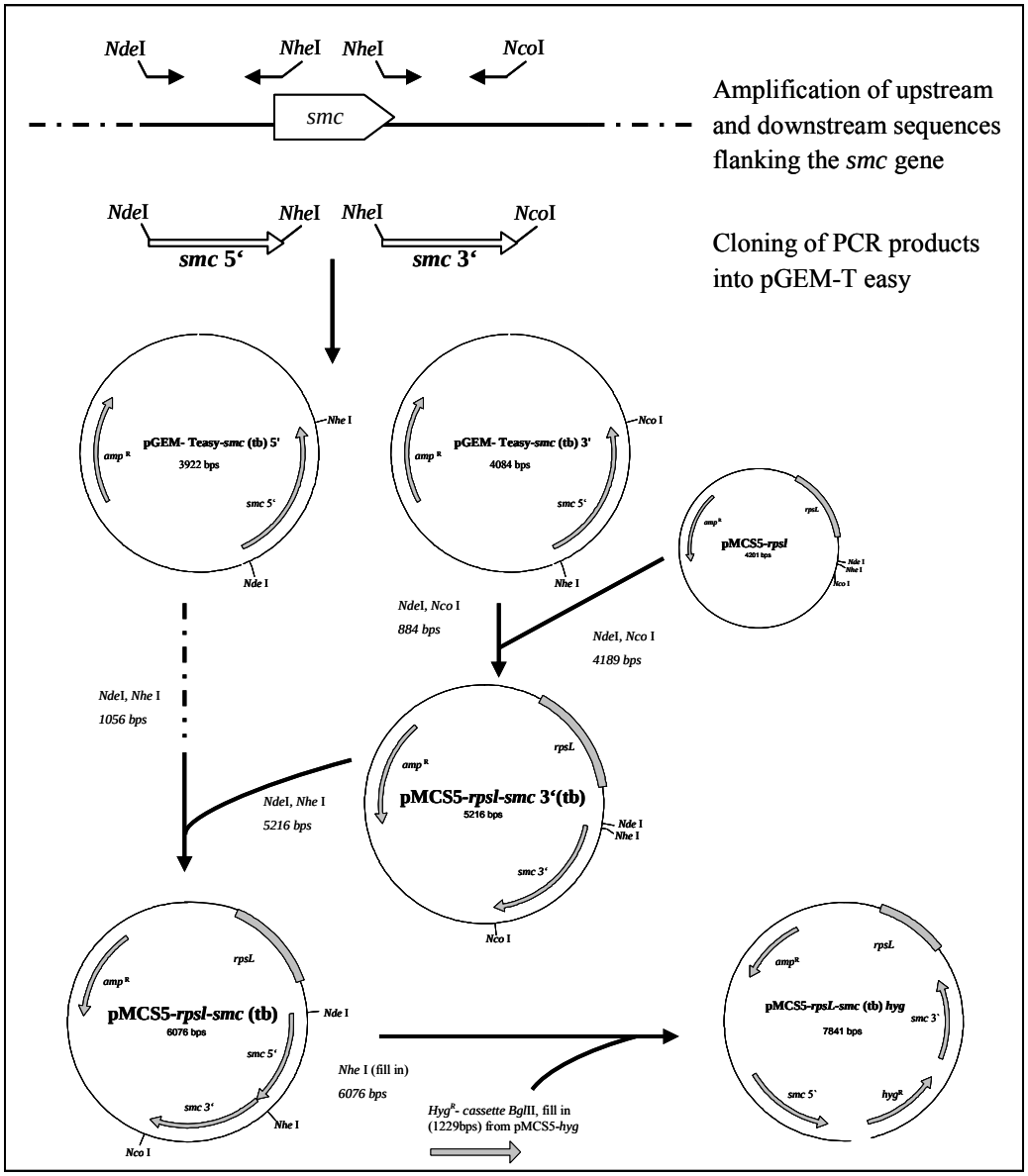


Figure 6: Construction of the knockout vector for the generation of an *M. tuberculosis smc* mutant

3.11 Assays

3.11.1 Spontaneous mutation frequencies

Mutation frequencies were determined by fluctuation experiments. Briefly, at least six parallel cultures of each strain were grown until late log phase in 7H9 medium, with a viable cell number around 10^9/ml. Subsequently, cultures were diluted to 2 x 10^3 / ml and incubated for 2 days at 37°C. 100 µl of each culture were plated on freshly prepared 7H10 agar plates containing rifampicin and serial dilutions were plated on non-selective medium. After 4 days of incubation at 37°C, colony forming units (CFU) were determined. The number of CFUs obtained on agar plates containing rifampicin was divided by the number of CFUs obtained on nonselective medium. For calculation of the mutation frequencies, the median of the ratio of cells that gained resistance against rifampicin was determined, as previously described (Lea and Coulson, 1949; Luria and Delbrück, 1943).

3.11.2 MIC and MBC

MICs were determined by E-test (AB BIODISK) according to the manufacturer's instructions. Determination of minimal bactericidal concentrations (MBCs) was performed in a microtiter plate format as described before (Pfister *et al.*, 2005). In brief, freshly grown cultures were diluted to an absorbance (A600) of 0.01 in 7H9 medium and incubated for 72 h at 37°C in the presence of 2-fold serial dilutions of rifampicin. The rifampicin stock solution (2 mg/ml) was made in DMSO. Aliquots from those wells that showed growth inhibition were plated on drug-free solid agar and incubated at 37°C for a further 72 h. The MBC is defined as the minimal drug concentration that kills >99.9 % of the bacterial inoculum.

3.11.3 Survival after exposure to UV light

UV sensitivity of strains was determined by plating dilutions (in triplicate) on solid medium and irradiation of open plates in a Stratalinker 2400 (Stratagene, 254nm) at 0–160 mJ/cm^2. CFUs were determined after 3-4 days incubation at 37°C and compared to the untreated control.

3.11.4 Survival after treatment with DNA damaging agents

DNA damaging agents (final concentration)

- Acidified sodium nitrite (pH 5.4, 3mM)
- *tert*-butylhydroperoxide (250µM)
- Mitomycin C (0.02 µg/ml)

25 ml cultures of each strain were grown in 7H9 medium until early midlog phase and serial dilutions were plated on LB agar for CFU determination. For determination of susceptibility to $NaNO_2$, cultures were centrifuged and resuspended in 7H9 medium acidified with HCl (pH 5.4). Subsequently, cultures were split into six 3 ml aliquots. Freshly prepared compounds (*tert*-butylhydroperoxide 250 µM, $NaNO_2$ 3 mM, mitomycin C 0,02 µg/ml) were added to three aliquots, the remaining aliquots served as untreated control. After incubation for 24 h at 37°C, serial dilutions were plated onto LB agar. CFUs were counted after 3 days growth at 37°C. Survival was calculated by the ratio of CFUs of the treated cultures compared to the CFUs of the untreated controls.

This protocol was also applied to *M. tuberculosis* strains, with the following modifications: Cultures were grown in 7H9 medium supplemented with OADC and exposed to the DNA damaging agent for 6 days. CFU were determined after 4-6 weeks growth at 37°C.

3.11.5 Susceptibility to ionizing radiation

Cultures were grown in 7H9 until early midlog phase. 1.5 ml of each strain were transferred into a 6 well cell culture titer plate and exposed to ionizing radiation (6.3 Gy/min.). 250 µl aliquots were taken after 10, 30 and 60 min of exposure. Serial dilutions were plated on LB agar plates. CFUs obtained after treatment were compared to the untreated control.

3.11.6 Gene conversion frequencies

M. smegmatis strains were transformed with an integrative vector carrying mutated *rrnA* gene fragments. The mutated *rrnA* gene fragments confer drug resistance (apramycin for the 16S rRNA 1491 mutation, clarithromycin for the 23S rRNA 2058/2059 mutation) upon recombination with the chromosomal wildtype *rrnA* gene (see section 4.3.5 and Table 3). For the positive selection of primary transformants, the vector encoded kanamycin resistance cassette was used. Following growth on medium containing kanamycin for three days, transformants were picked and grown in 4 ml 7H9 medium for another three days. Gene

conversion frequencies were assessed by fluctuation experiments. Briefly, for each construct and strain, at least ten independent transformants were picked and grown. Aliquots of these precultures were taken to inoculate 4 ml 7H9 medium with approximately $5x10^5$ cells. After two days of growth, serial dilutions were plated on permissive medium and on selective medium containing either apramycin or clarithromycin. The relative marker integration frequencies were determined by the ratio of cells which gained antibiotic resistance compared to the number of cells obtained on permissive medium.

3.12 Mice infection experiments

BALB/c and nude mice (6 to 8 weeks old) were obtained from the breeding facility at the National Institute for Medical Research (Mill Hill, United Kingdom). *M. tuberculosis* strains were grown in Dubos broth. Logarithmically growing cultures were diluted in saline to an OD of 0.8; 0.2 ml (approximately 10^6 CFU) were injected into the tail vein. Mice were sacrificed according to ethical guidelines at various times (three mice per strain for each time point), and the spleens and lungs were removed, weighed, and homogenized. The suspensions were serially diluted in saline and plated on 7H10 agar supplemented with OADC. CFU were determined after 3 weeks incubation at 37°C and calculated as CFU per organ. These experiments were performed in collaboration with Elaine O. Davis (National Institute for Medical Research, London, UK).

4 Results

4.1 NER in mycobacteria

Genome analyses of *M. tuberculosis* and *M. smegmatis* indicate that both species have a full set of NER genes. The open reading frame MSMEG3816 has been annotated as excinuclease ABC subunit B (*uvrB*) in the genome of *M. smegmatis*. MSMEG3816 encodes a protein of 719 amino acids. It displays a high degree of homology to *M. tuberculosis uvrB* (Rv1633).
M. tuberculosis uvrB encodes a protein of 698 amino acids, with 83% identity (87% positives) to *M. smegmatis uvrB*.
The open reading frame MSMEG5534 in the genome of *M. smegmatis* has been annotated as ATP-dependent DNA helicase and has been named *pcrA* in the TIGR genome data base. Since the encoded protein (783 amino acids) has a high degree of homology (80% identity / 87% positives) to *M. tuberculosis uvrD* (Rv0949), we propose that MSMEG5534 represents the *M. smegmatis uvrD* homologue. In this context it is important to note that in other Gram-positive bacteria like *B. subtilis uvrD* homologues are named *pcrA* as well.

4.2 Generation of *M. smegmatis* NER mutants

To study the role of the NER pathway in mycobacteria, *M. smegmatis* mutants deficient in *uvrB* and in *uvrD* were constructed.
For targeted gene inactivation an unmarked deletion mutagenesis strategy employing a suicidal vector that harbours a counter-selectable marker was used as described in section 3.10. Disruption of *uvrD* was verified by Southern blot analysis using a DNA probe specific for the 5' flanking region of the *uvrD* open reading frame. This analysis revealed a 10.7 kbp hybridization signal in the mutant strain compared to a 2.0 kbp fragment in the parental strain. The difference in size is due to the loss of two restriction sites that are located in the deleted region (Figure 7 a).
Disruption of *uvrB* was verified by Southern blot analysis using an *uvrB*-specific probe. This analysis revealed a 3.9 kbp hybridization signal in the mutant strain compared to a 5.6 kbp fragment in the parental strain. The difference in size is due to deletion of the genomic *uvrB* region (Figure 7 b).

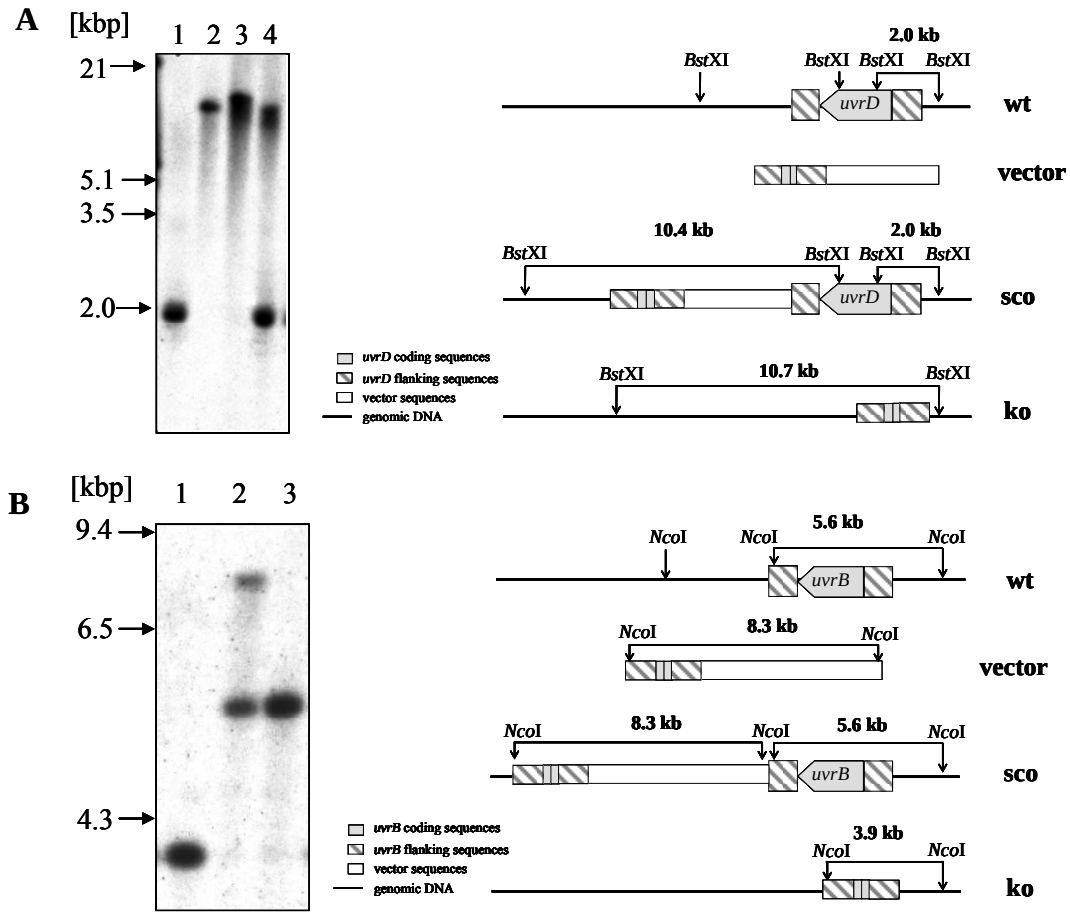


Figure 7: Generation of *M. smegmatis* NER single mutants

A Disruption of *M. smegmatis uvrD* Left: Southern blot analysis. Genomic DNA from *M. smegmatis* wildtype (lane 1), *M. smegmatis uvrD* mutant (lanes 2-3) and *M. smegmatis uvrD* sco (lane 4) was digested with *Bst*XI and probed with a 653 bp *Mlu*I/*BamH*I DNA fragment containing 5' flanking regions of the *uvrD* gene. The presence of a single 10.7 kbp fragment instead of a 2.0 kbp fragment as seen in the parental strain demonstrates successful deletion of *uvrD* coding sequences. Right: Schematic illustration of the *uvrD* locus and the Southern blot strategy. Shown are the genomic organisation of wildtype (wt), the knockout vector that contains the deletion allele (vector), the single cross-over genotype (sco) and the mutated genomic *uvrD* region in the knockout mutant (ko). Fragments detected by the probe specific for the 5' flanking region are shown in bold letters.

B Disruption of *M. smegmatis uvrB* Left: Southern blot analysis. Genomic DNA from *M. smegmatis uvrB* mutant (lane 1), *M. smegmatis uvrB* sco (lane 2) and *M. smegmatis* wildtype (lane 3) was digested with *Nco*I and probed with a 600 bp *Nde*I/*Bsm*I *uvrB* gene fragment. The presence of a single 3.9 kbp fragment instead of a 5.6 kbp fragment as seen in the parental strain demonstrates successful deletion of *uvrB* coding sequences. Right: Schematic illustration of the *uvrB* locus and the Southern blot strategy. Shown are the genomic organisation of wildtype (wt), the knockout vector that contains the deletion allele (vector), the single cross-over genotype (sco) and the mutated genomic *uvrB* region in the knockout mutant (ko). Fragments detected by the probe specific for the *uvrB* gene are shown in bold letters.

A *M. smegmatis* double mutant deficient in *uvrB* and *uvrD* was obtained after transformation of the *uvrD* mutant strain with the *uvrB* knockout vector. Disruption of *uvrB* was confirmed by Southern blot analyses using hybridization probes specific for *uvrD* or *uvrB*, respectively. (Figure 8). The schematic illustration of the *uvrD* and *uvrB* gene loci and the Southern blot strategies are given in Figure 7.

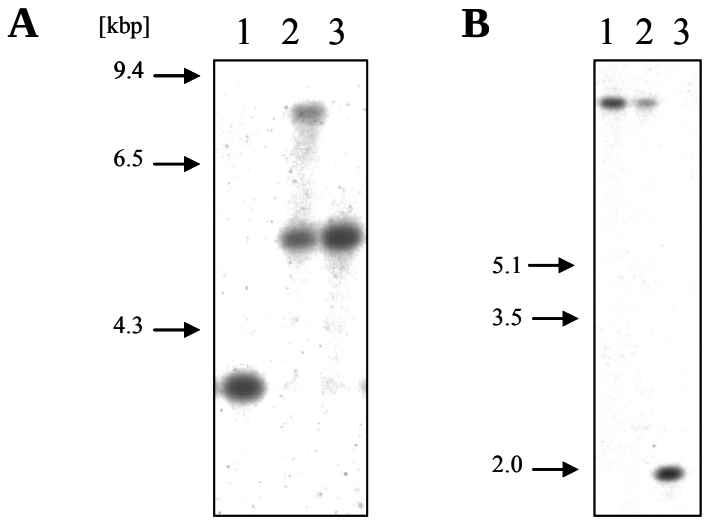


Figure 8: Generation of *M. smegmatis uvrD/ uvrB* double mutant

(A) Genomic DNA from *M. smegmatis uvrD / uvrB* double mutant (lane 1), *M. smegmatis uvrD* mutant / *uvrB* sco (lane 2) and *M. smegmatis* wildtype (lane 3) was digested with *Nco*I and probed with an oligonucleotide specific for *uvrB*. The presence of a single 3.9 kbp fragment instead of a 5.6 kbp fragment as seen in the parental strain demonstrates successful deletion of *uvrB* coding sequences.

(B) Genomic DNA from *M. smegmatis uvrD / uvrB* double mutant (lane 1), *M. smegmatis uvrD* mutant / *uvrB* sco (lane 2) and *M. smegmatis* wildtype (lane 3) was digested with *BstX*I and probed with an oligonucleotide specific for u*vrD*. The presence of a single 10.7 kbp fragment instead of a 2.0 kbp fragment as seen in the parental strain demonstrates successful deletion of *uvrD* coding sequences.

4.3 Characterisation of *M. smegmatis* NER mutant strains

4.3.1 *In vitro* growth

The *in vitro* growth characteristics of *M. smegmatis* wildtype and the NER mutants were assessed in 7H9 medium. For comparison, a *M. smegmatis recA* mutant (Frischkorn *et al.*, 1998) was included to the studies. Doubling times of the *uvrB* and *uvrD* mutant strains were indistinguishable from the parental strain. Growth of the *uvrB* / *uvrD* double mutant was considerably impaired, with a twofold increase in generation time compared to the wildtype (Figure 9). With respect to colony morphology, the mutants did not show any difference to the wildtype. When grown on solid media, the double mutant strain showed visible colonies about 20 hours later than the wildtype, confirming the general growth defect of the double mutant strain (data not shown).

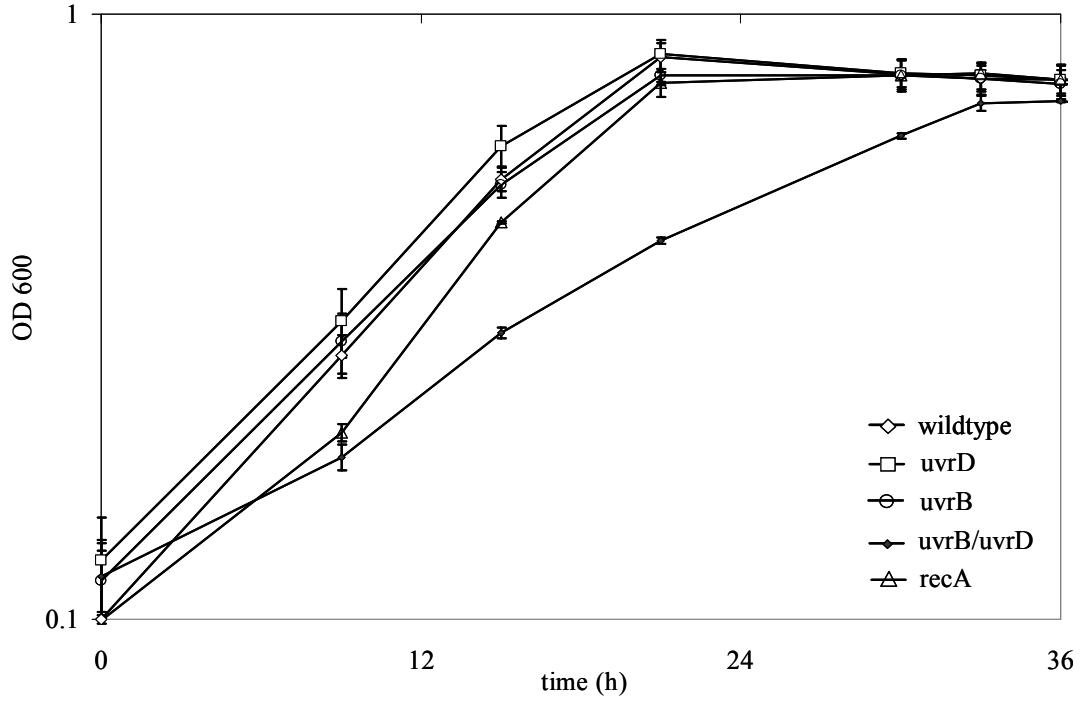


Figure 9: Growth analysis of *M. smegmatis* strains

Doubling times: wildtype 3.7 h +/- 0.1, *uvrD* mutant 3.5 h +/- 0.2, *uvrB* mutant 3.4 h +/- 0.2, *uvrB* / *uvrD* mutant 7.6 h +/- 0.4, *recA* mutant 3.9 h +/- 0.2.

4.3.2 Spontaneous mutation frequencies

Rifampicin has a single molecular target, the ß subunit of RNA polymerase encoded by the *rpoB* gene. Various point mutations in *rpoB* may confer high-level resistance. The frequency at which bacteria generate rifampicin-resistant mutants is widely used to assess the spontaneous mutation frequency. The calculation is based on the model for the distribution of mutant clones originally described by Luria and Delbrück (1943) and extended by Lea and Coulson (1949).

The frequencies at which repair-deficient *M. smegmatis* mutants gain rifampicin resistance were investigated as described in section 3.11.1 and compared with those of the parental wildtype strain (Figure 10). The wildtype strain displayed a spontaneous mutation frequency of 3.8×10^{-8}. In comparison, deletion of *uvrD* resulted in an approximately threefold increased mutation frequency (9.1×10^{-8}, $p < 0.05$, n= 6, student's t-test) and deletion of *uvrB* resulted in a fivefold increased mutation frequency (1.9×10^{-7}, $p < 0.01$, n= 6, student's t-test). Inactivation of *recA* did not affect the mutation frequency (1.7×10^{-8} $p > 0.05$, n= 6, student's t-test). The mutation frequency of the double mutant could not be assessed with this approach, as the *uvrB / uvrD* mutant strain was unable to grow on the selective plates used for recovery of the spontaneous drug resistant mutants (containing 175 µg/ml rifampicin). Considering the general growth defect of this strain, we assumed that the *uvrB / uvrD* mutant is incapable of acquiring high-level resistance against rifampicin. Determining the minimal inhibitory concentration (MIC) of rifampicin indicated that the *uvrB / uvrD* mutant displayed an eightfold decreased MIC (4 µg/ml compared to 32 µg/ml in the wildtype and the *uvrB* mutant strain and 16 µg/ml in the *uvrD* mutant strain, Table 4). Analysis of the minimal bactericidal concentration (MBC) revealed that no viable clones of the double mutant could be recovered following incubation at rifampicin concentrations higher than 130 µg /ml (for the wildtype strain and the *uvrB* and *uvrD* single mutant strains the MBC was 500 µg /ml, Table 4). Given the decreased MBC of rifampicin, we assessed the spontaneous mutation frequency of the *uvrB* / *uvrD* double mutant on plates containing 75 µg/ml rifampicin. Using this approach, the *uvrB* / *uvrD* double mutant strain was found to display an approximately 20 fold increased spontaneous mutation frequency (7.2×10^{-7}, $p < 0.01$, n= 6, student's t-test) compared to the wildtype strain (Figure 10).

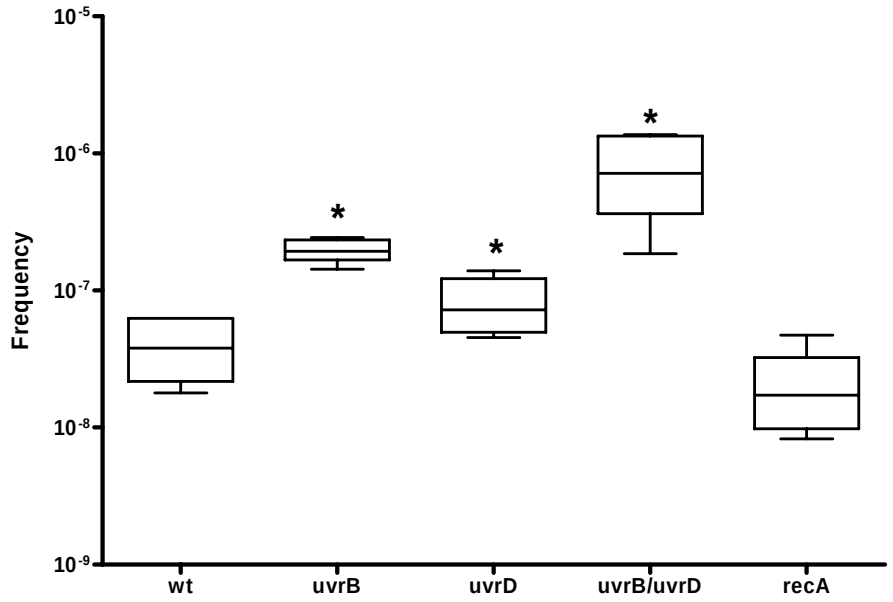


Figure 10: Spontaneous mutation frequencies of *M. smegmatis* wildtype (wt) and mutant strains

Shown are the mean values (horizontal lines), 95% confidence interval (rectangular boxes) and outliers (vertical lines) of mutation frequencies calculated from the ratio of spontaneous rifampicin resistant cells obtained in six independent cultures of each strain. * Significant increase of mutation frequency compared to the wildtype statistically confirmed (student's t-test $p<0.05$). Abbreviations: wt = *M. smegmatis* wildtype strain; *uvrB*, *uvrD*, *uvrB*/*uvrD* = *M. smegmatis* NER mutant strains, *recA* = *M. smegmatis recA* mutant strain.

Table 4: MIC and MBC to rifampicin of *M. smegmatis* NER mutant strains

strain	MIC (µg / ml)	MBC (µg / ml)
wildtype	32	500
uvrB	32	500
uvrD	16	500
uvrB / uvrD	4	130

4.3.3 Survival after treatment with DNA damaging agents

To investigate whether absence of NER in *M. smegmatis* affects DNA repair proficiency, mutants were challenged with DNA-damaging agents in a survival assay. Different DNA damaging compounds were chosen to cover various types of DNA damage.

i.) *Tert*-butyl-hydroperoxide (TBH) generates reactive oxygen intermediates (ROI) which can attack the bases and the sugar-phosphate backbone of the DNA, leading to strand breaks and base adducts (Halliwell and Aruoma, 1991). Treatment with TBH possibly also results in formation of alkoxyradicals and methylation reactions of the DNA, thus leading to alkylation damage (Hix *et al.*, 1995). The *M. smegmatis* NER mutants were all significantly more susceptible towards TBH compared to the wildtype strain (Figure 11 a). The *uvrB*, *uvrD* and the *uvrB / D* double mutant displayed similar susceptibilities (student's t-test: $p > 0.1$). TBH susceptibility of the *M. smegmatis recA* mutant was only slightly, yet significantly increased compared to the wildtype (42 % surviving cells of the *recA* mutant compared to 77% surviving cells of the wildtype; $p < 0.05$ student's t-test). These findings indicate that NER is more important than recombinational repair in defence against ROI and alkylation damage in mycobacteria.

ii.) Acidified sodium nitrite generates reactive nitrogen (RNI) intermediates, leading to oxidative damage of the DNA as well as inducing nitration, nitrosation and deamination reactions (Burney *et al.*, 1999). The *M. smegmatis recA* mutant displayed a similar susceptibility to RNI as the wildtype (12% and 15 % surviving cells; $p > 0.08$ student's t-test, Figure 11 b). As with ROI, the *M. smegmatis* NER mutants displayed significantly increased susceptibilities to acidified sodium nitrite. The *uvrD* and the *uvrB / uvrD* double mutant were considerably more susceptible than the *uvrB* single mutant (Figure 11 b), implying a further role of *uvrD* in the repair of RNI induced DNA damage.

iii.) Exposure to mitomycin C (MMC) results in alkylation and interstrand crosslinks (Iyer and Szybalski, 1963; Kumar *et al.*, 1992), which are mainly repaired by NER in *E. coli* (Van Houten *et al.*, 2005). The *M. smegmatis* NER mutants diplayed different susceptibilities towards treatment with MMC. While susceptibilities of the *recA* and the *uvrD* mutants were substantially increased compared to the wildtype, the *uvrB* and the *uvrB / D* double mutants were extremely susceptible towards MMC (Figure 11 c).

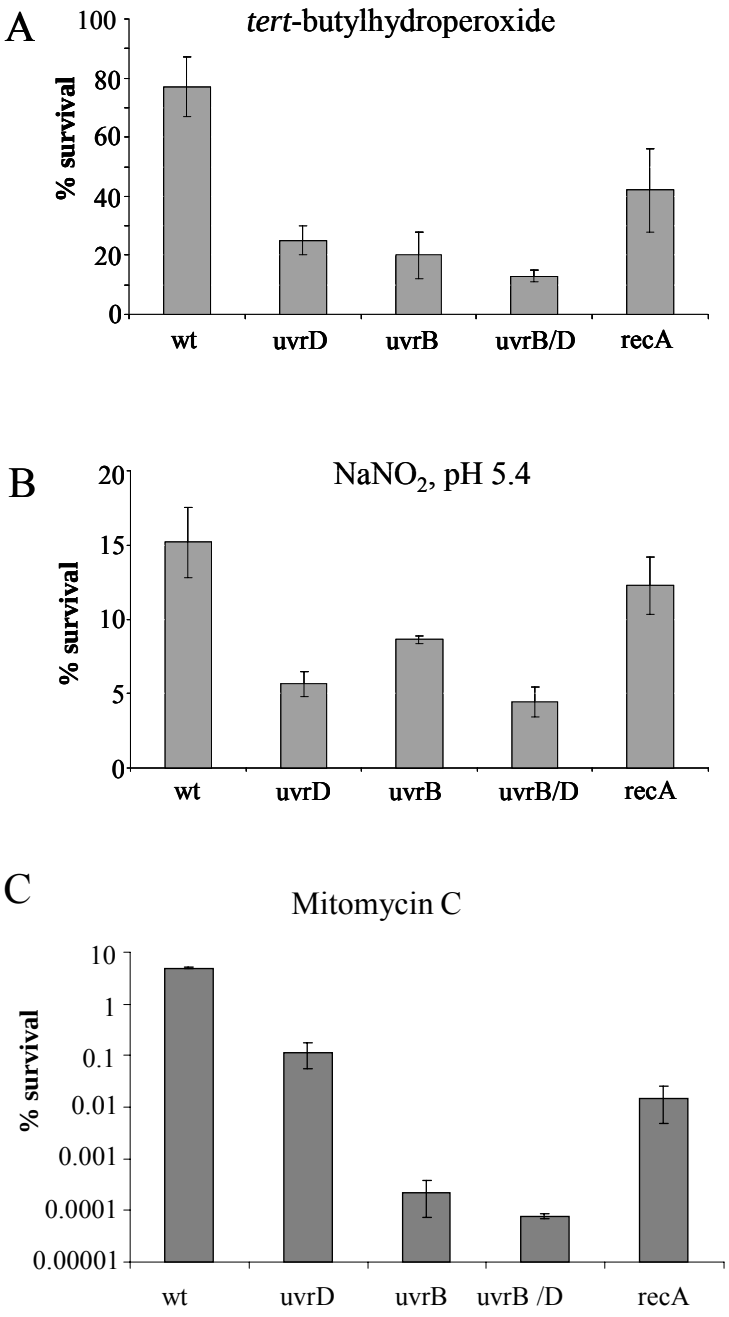


Figure 11: Survival of *M. smegmatis* strains following treatment with different DNA damaging agents

(A) *tert*-butyl-hydroperoxide (250 µM); **(B)** acidified $NaNO_2$ (pH 5.4, 3mM); **(C)** mitomycin C (0.02 µg/ml). Survival was calculated from the ratio of cells that survived treatment with the DNA damaging agent in proportion to CFU of the untreated control. Strains were incubated with the DNA damaging agent for 24 h. Shown are the mean values of one representative experiment performed with three independent cultures each. Standard deviations from the mean are given in the error bars. Abbreviations: wt = *M. smegmatis* wildtype strain; *uvrB*, *uvrD*, *uvrB/D* = *M. smegmatis* NER mutant strains, *recA* = *M. smegmatis recA* mutant strain.

4.3.4 Survival after treatment with short wavelength UV light

NER mutants in *E. coli* display an increased UV-susceptibility and NER is responsible for the repair of bulky adducts resulting from UV-light exposure. Hence, the impact of short wavelength UV light (UV-C) to survival of the *M. smegmatis* NER mutants was studied. Increasing UV-C doses from 0 mJ / cm^2 to 160 mJ / cm^2 were applied and *M. smegmatis* wildtype and mutant strains irradiated as described in section 3.11.2.

The *uvrB* and *uvrD* mutant strains as well as the *recA* mutant strain were considerably more susceptible than the wildtype strain. Interestingly, combined deletion of *uvrB* and *uvrD* had an additive effect; cells were unable to survive doses in excess of 80 mJ / cm^2 (Figure 12).

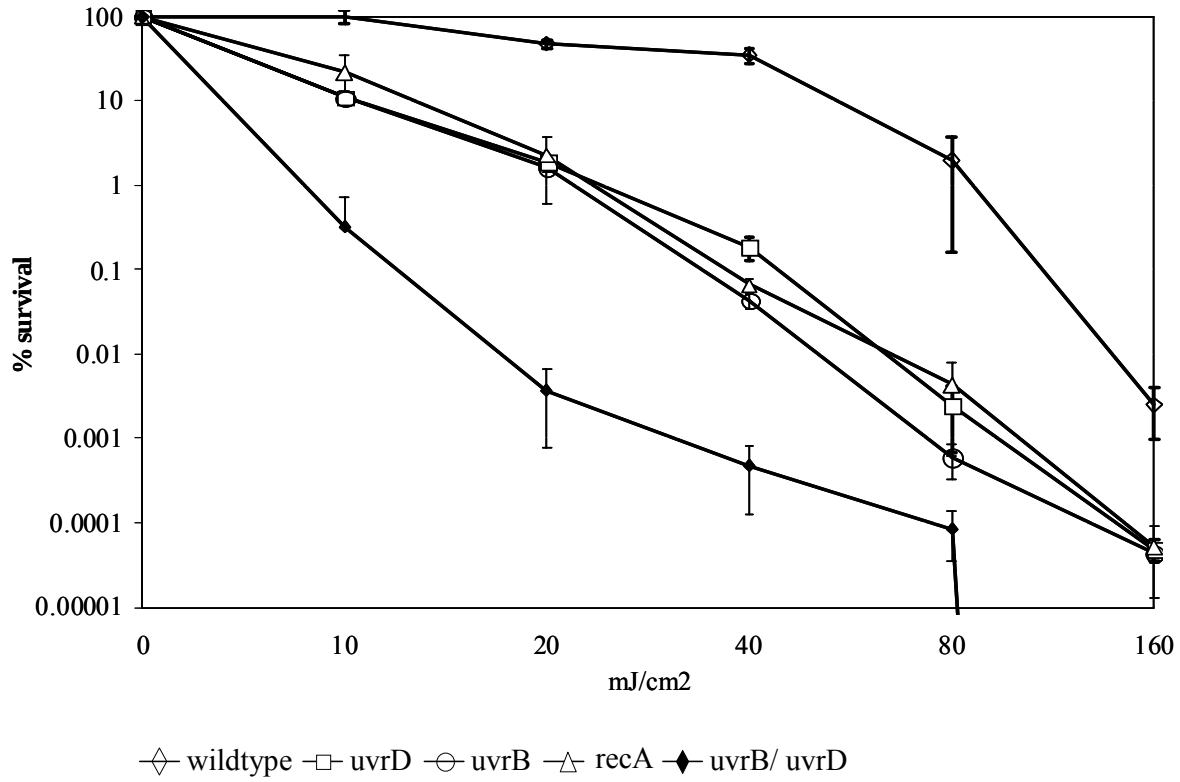


Figure 12: Survival of *M. smegmatis* strains following irradiation with UV-C.

Survival was calculated from the ratio of cells that survived treatment with UV-C in proportion to the untreated control. Shown are the mean values of one representative experiment performed with three independent cultures each. Standard deviations of the means are given in the error bars. Ratio of treated to untreated cells is plotted in logarithmic scale. The *uvrB / D* mutant was unable to survive doses in excess of 80 mJ / cm^2.

4.3.5 Gene conversion assays

Base-base mispairing occurs in duplex DNA in the mode of purine-purine (G/G, A/A, G/A), purine-pyrimidine (G/T, A/C) or pyrimidine-pyrimidine (C/C, T/T, T/C) mismatches. With the exception of C/C, all base-base mispairings are subject to correction by the mismatch repair system (Marra and Schär, 1999). As mycobacteria are devoid of this otherwise highly conserved repair system, the effects of MMR deficiency and possible compensating roles of NER in postreplicative repair were examined. Towards this end, a plasmid transformation-based assay was applied. Upon integration at the *attB* site of the mycobacterial genome, intermolecular recombination between the chromosomal *M. smegmatis rrnA* gene and the plasmid-borne *rrnA* gene fragment carrying a specific resistance mutation leads to antibiotic resistance. Intrachromosomal recombination occurs via an intermediate with a non-Watson-Crick base pairing. In repair proficient cells, the intermediate is recognised and rejected, resulting in decreased gene conversion frequencies. In case of repair deficiency, the intermediate will not be recognised as abnormal, leading to mutation of the chromosomal *rrnA* operon following replication and postmitotic segregation in 50 % of the progeny, with consequent gain in resistance to the relevant antibiotic (Figure 13). The assay provided the opportunity to analyse recognition of different mismatches, e.g. purine-pyrimidine (G/T, A/C) or pyrimidine-pyrimidine (C/C, C/T, T/C) mismatches. The *M. smegmatis* wildtype chromosome harbours a second *rrn* operon (named *rrnB*) in addition to the *rrnA* operon, which precludes proper measurement of gene conversion frequencies using the assay described above. Hence, the NER mutants were constructed in a Δ *rrnB* background (Sander *et al.*, 1996).

M. smegmatis rrnB NER mutant strains were transformed with an integrative vector carrying a mutated *rrnA* gene fragment. The relative recombination frequencies using different *rrnA* fragments were analysed. Each of the *rrnA* fragments comprises a specific resistance conferring mutation which leads to a defined base-pairing error in the intermediate upon recombination (Table 5). The relative recombination frequency was assessed by calculating the ratio of cells which gained antibiotic resistance compared to the number of cells obtained on medium without antibiotic selection. The recombination frequencies were determined according to Lea and Coulson (1949) from fluctuation experiments with at least 10 independent cultures for each strain and *rrnA* fragment.

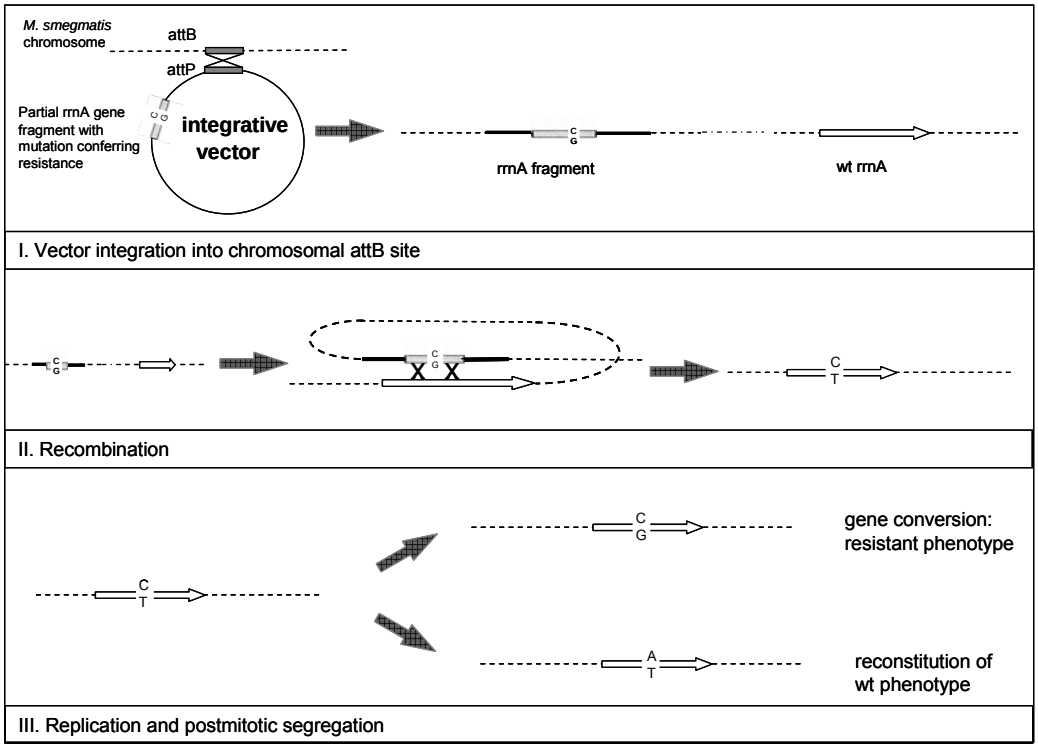


Figure 13: Schematic overview of the assay applied in this study

Upon integration of the vector at the chromosomal attachment site (*attB*), intrachromosomal recombination leads to an intermediate with a non-Watson-Crick base pairing. In case of mismatch repair deficiency, the intermediate is not rejected. After replication and postmitotic segregation, 50 % of the progeny gain resistance to an antibiotic.

Table 5: Recombination substrates applied in this study

position	resistance mutation	resulting mismatch	resistance
23S rRNA 2058	A → C	C/T	Clarithromycin
23S rRNA 2058	A → G	G/T	Clarithromycin
23S rRNA 2059	A → C	C/T	Clarithromycin
23S rRNA 2059	A → G	G/T	Clarithromycin
16S rRNA 1491	G → A	A/C	Apramycin
16S rRNA 1491	G → T	T/C	Apramycin
16S rRNA 1491	G → C	C/C	Apramycin

For each of the different *rrnA* fragments, specific gene conversion frequencies were obtained. Comparison of marker integration frequencies observed in the wildtype revealed that the highest gene conversion frequencies (in the range of 10^{-4} to 10^{-5}) were found with an A → C mutation leading to a C/T mismatch (Figure 14a). This mismatch was studied

making use of two different constructs with the A → C mutation located at 23S rRNA position 2058, or 2059, respectively. Deletion of *uvrD* resulted in significantly (approximately tenfold) increased gene conversion frequencies in comparison to the wildtype. In contrast, gene conversion frequencies in the *uvrB* mutant were virtually at wildtype control level (student's t-test, n=10, p>0.05). Interestingly, combined deletion of *uvrB* and *uvrD* resulted in significantly elevated marker integration frequencies (about 5 fold [A → C mutation located at 23S rRNA position 2058] or 20 fold [A → C mutation located at 23S rRNA position 2059]) compared to the frequencies obtained in the *uvrD* single mutant (Figure 14 a). Analysis of marker integration frequencies with an *rrnA* fragment carrying an A → G mutation which leads to a G/T mismatch, gave similar results (Figure 14 a). Again, two different constructs carrying the mutation at 23S rRNA position 2058, or 2059, were used. In general, marker integration frequencies were lower with the A → G mutation than with the A → C mutation, yielding frequencies in the range of 10^{-5} and 10^{-6} in the *M. smegmatis* wildtype. Gene conversion frequencies were considerably (about tenfold) elevated in the *uvrD* mutant, but only marginally (3 to 4 fold) increased in the *uvrB* mutant. A synergism of combined absence of *uvrB* and *uvrD* was observed, as the double mutant displayed marker integration frequencies about 20 fold higher than those of the *uvrD* single mutant (Figure 14 a).

Recognition of A/C and C/C mismatches was determined with *rrnA* fragments carrying a G → A mutation or a G → C mutation at 16S rRNA position 1491. Gene conversion frequencies of the wildtype were in the range of 1 x 10^{-6}. Gene conversion frequencies of the *uvrB* mutant, the *uvrD* mutant and the *uvrB* / *uvrD* double mutant were only marginally (about fourfold) increased. Thus, no synergism of combined deletion of *uvrB* and *uvrD* was observed for A/C and C/C mismatch recognition (Figure 14 b). Analysis of marker integration frequencies using an *rrnA* fragment carrying a G → T mutation provided insights into recognition of T/C mismatches. The *uvrD* mutant strain and the *uvrB* / *uvrD* mutant strain yielded approximately tenfold increased gene conversion frequencies, whilst the *uvrB* mutant strain yielded fourfold increased frequencies (Figure 14 b).

In summary, different gene conversion frequencies depending on the mismatch studied were observed. The NER mutants displayed higher marker integration frequencies compared to the wildtype. While the *uvrB* mutant showed only slightly increased gene conversion frequencies, high marker integration frequencies were observed with the *uvrD* mutant. The *uvrB / uvrD* mutant displayed the highest gene conversion frequencies, however, only with some of the mismatches studied.

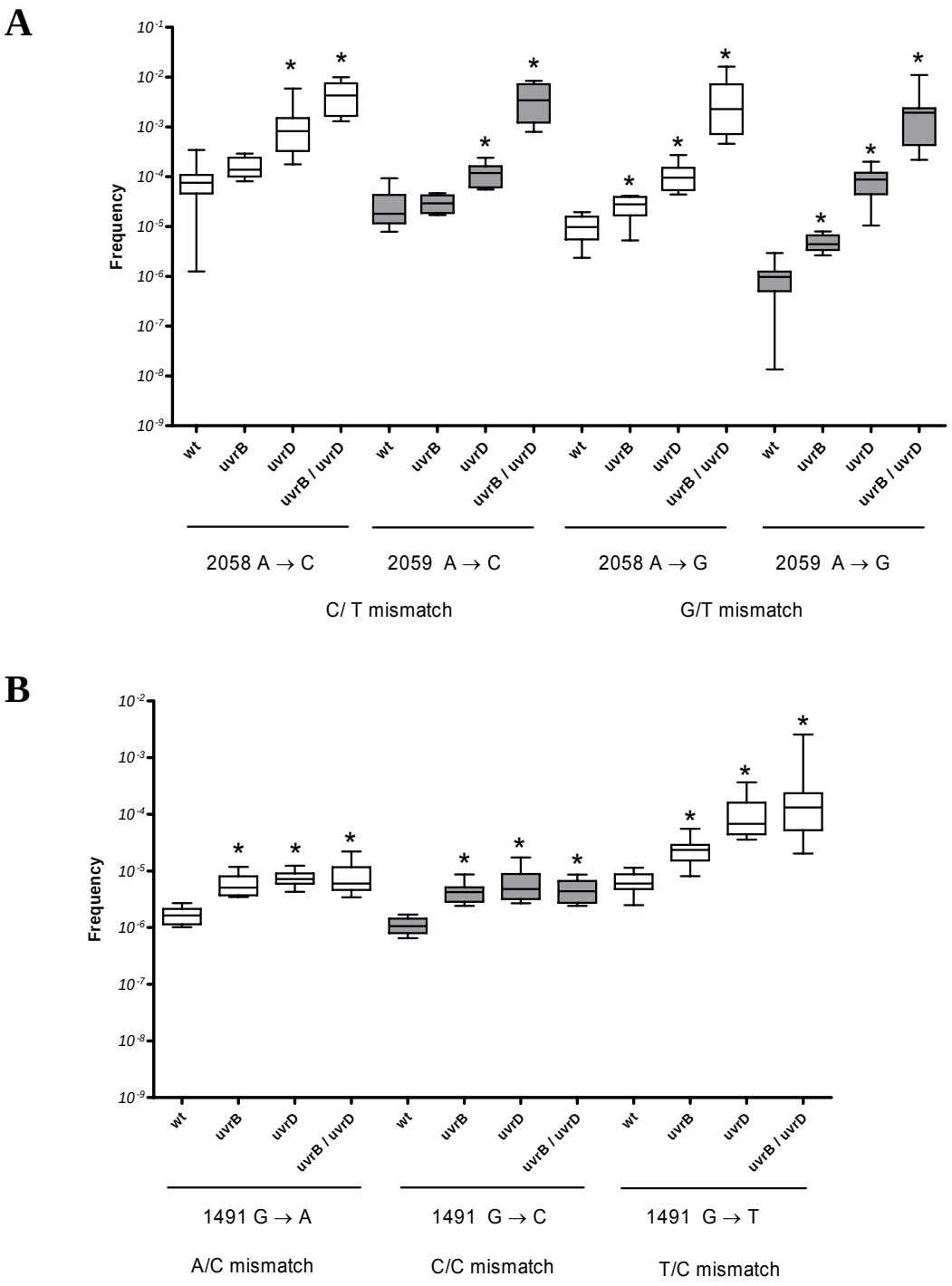


Figure 14: Relative gene conversion frequencies obtained with *rrnA* fragments resulting in C/T or G/ T mismatches (A) and *rrnA* fragments resulting in A/C, C/C or T/C mismatches (B)

Shown are the median values (horizontal lines), 95% confidence interval (rectangular boxes) and outliers (vertical lines) of recombination frequencies calculated from the number of cells that gained antibiotic resistance in ten independent cultures of each strain. * Significant increase of gene conversion frequencies compared to the wildtype (student's t-test, $p<0.05$). Abbreviations: wt = *M. smegmatis* wildtype strain; *uvrB*, *uvrD*, *uvrB*/*uvrD* = *M. smegmatis* NER mutant strains.

4.4 Disruption of NER DNA repair pathway in *M. tuberculosis*

To assess the role of NER in the pathogenesis of *Mycobacterium tuberculosis* infection, mutants deficient in *uvrA* and in *uvrD* were generated, as well as a double mutant deficient in both proteins. The mutants were constructed by transformation with suicide plasmids using the *rpsL* counterselection strategy described in section 3.10.

Disruption of *uvrD* was confirmed by Southern blot analysis making use of a probe comprising a 2411 bp *Hpa*I / *Sac*I *uvrD* gene fragment. This analysis revealed a hybridization signal in the mutant strain at 5.3 kbp compared to a signal at 3.7 kbp in the parental strain. The shift in fragment size in the *uvrD* knockout strain results from replacement of a 1.4 kbp *uvrD* gene fragment by a 3 kbp gentamicin resistance cassette. The mutant strain was complemented by introducing plasmid pMV361-*uvrD* expressing the wildtype *uvrD* allele, resulting in *M. tuberculosis uvrD* compl. (Figure 15).

Disruption of *uvrA* was verified by Southern blot analysis using a probe specific for the 5' flanking region of the *uvrA* open reading frame. This analysis revealed a 5.6 kbp hybridization signal in the mutant strain and a 4.3 kbp signal in the parental strain. The difference in size is due to loss of a restriction site after replacement of a 1.3 kbp *uvrA* gene fragment with a 1.2 kbp kanamycin resistance cassette (Figure 16).

A *M. tuberculosis* double mutant deficient in *uvrA* and *uvrD* was obtained after transformation of the *uvrA* mutant strain with the *uvrD* knockout vector. Disruption of *uvrD* was verified by two Southern blot analyses using hybridization probes specific for *uvrD* and *uvrA*, respectively. The double mutant exhibited the hybridization patterns expected for disruption of *uvrD* and *uvrA*, respectively (Figure 15, Figure 16). Complementation of the double mutant is ongoing.

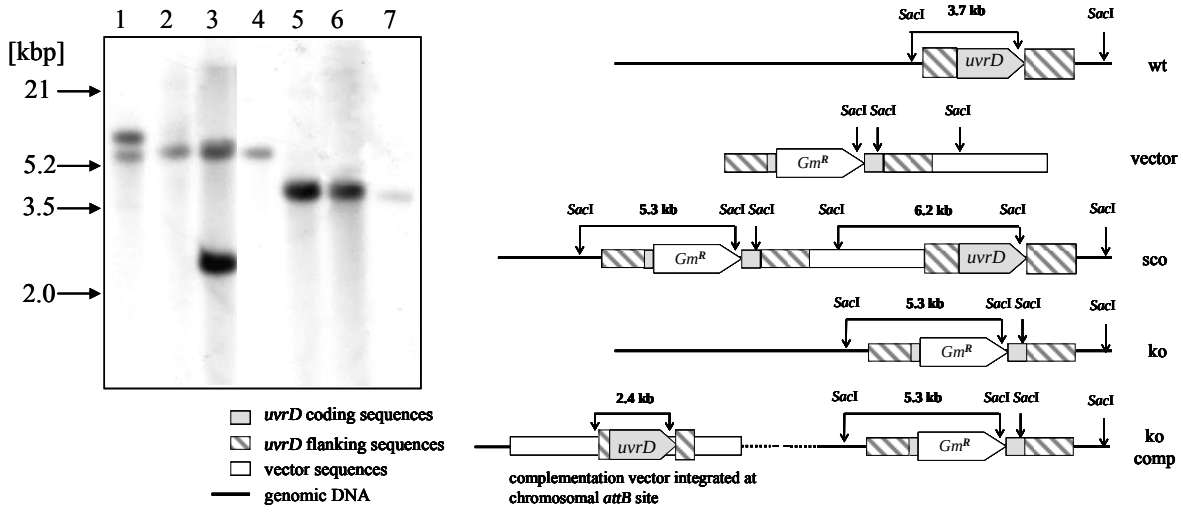


Figure 15: Generation of *M. tuberculosis uvrD* and *uvrA/ D* mutants

Left panel: Southern blot hybridised with a *uvrD*-specific probe. Genomic DNA from *M. tuberculosis uvrD* sco (lane 1), *uvrD* mutant (lane 2), complemented *uvrD* mutant (lane 3), *uvrA* / *uvrD* mutant (lane 4), *uvrA* sco (lane 5), *uvrA* mutant (lane 6) and wildtype (lane 7) was digested with *Sac*I and probed with a 2411 bp *Hpa*I / *Sac*I *uvrD* gene fragment. The presence of a single 5.3 kbp fragment instead of a 3.7 kpb fragment demonstrates successful deletion of *uvrD* coding sequences. Complementation of *uvrD* is indicated by an additional hybridization signal at 2.4 kbp. The *uvrA* single cross-over (sco) and the *uvrA* mutant exhibit a hybridization pattern identical to the wildtype pattern.

Right panel: Schematic illustration of the *uvrD* locus and the Southern blot strategy. Shown are the genomic organisation of wildtype (wt), the knockout vector that contains the deletion allele *uvrD::gm* (vector), the single cross-over genotype (sco), the mutated genomic *uvrD* region in the knockout mutant (ko) and its complementation at the *attB* site (ko comp). Fragments detected by the probe specific for the 5' flanking region are shown in bold letters.

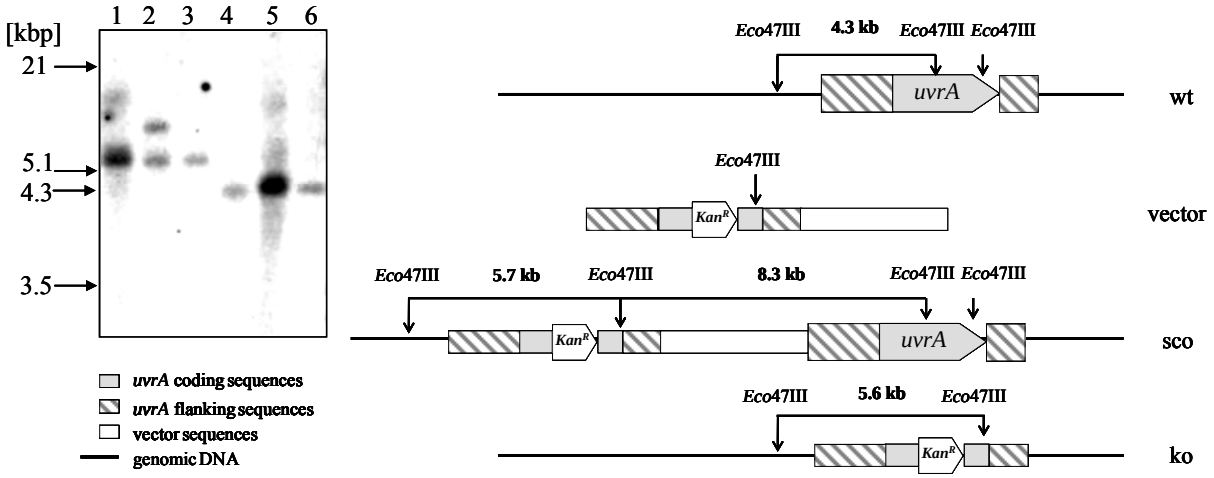


Figure 16: Generation of *M. tuberculosis uvrA* and *uvrA/ D* mutants

Left panel: Southern blot hybridised with a probe specific for the upstream flanking regions of the *uvrA* gene. Genomic DNA from *M. tuberculosis uvrA* mutant (lane 1), *uvrA* sco (lane 2), *uvrA / uvrD* mutant (lane 3), *uvrD* sco (lane 4), *uvrD* mutant (lane 5) and wildtype (lane 6) was digested with *Eco*47III. The presence of a single 5.6 kbp fragment instead of a 4.3 kbp fragment demonstrates successful deletion of *uvrA* coding sequences. The *uvrD* single cross-over (sco) and the *uvrD* mutant show a hybridization pattern identical to the wildtype pattern.

Right panel: Schematic illustration of the *uvrA* locus and the Southern blot strategy. Shown are the genomic organisation of wildtype (wt), the knockout vector that contains the deletion allele *uvrA::kan* (vector), the single cross-over genotype (sco) and the mutated genomic *uvrA* region in the knockout mutant (ko). Fragments detected by the probe specific for the 5' flanking region are shown in bold letters.

4.5 *In vitro* characterisation of *M. tuberculosis* NER mutant strains

4.5.1 *In vitro* growth

The NER mutants differed from the wildtype strain by remarkable changes in colony morphology as well as in colony size. The colonies of the NER mutant strains were significantly smaller than colonies of the wildtype strain. The most considerable decrease in colony size was found with the *uvrA / uvrD* mutant strain (Figure 17).

To analyse the phenotype of the *M. tuberculosis* NER mutant strains in more detail, the *in vitro* growth in 7H9 broth supplemented with OADC was studied. Compared to the wildtype strain, the *uvrA / uvrD* mutant displayed considerable growth retardation. Deletion of *uvrA* had a slight effect on growth rate, whereas the *uvrD* mutant strain and the complemented *uvrD* mutant strain displayed growth rates virtually identical to the growth rate of the wildtype strain (Figure 18).

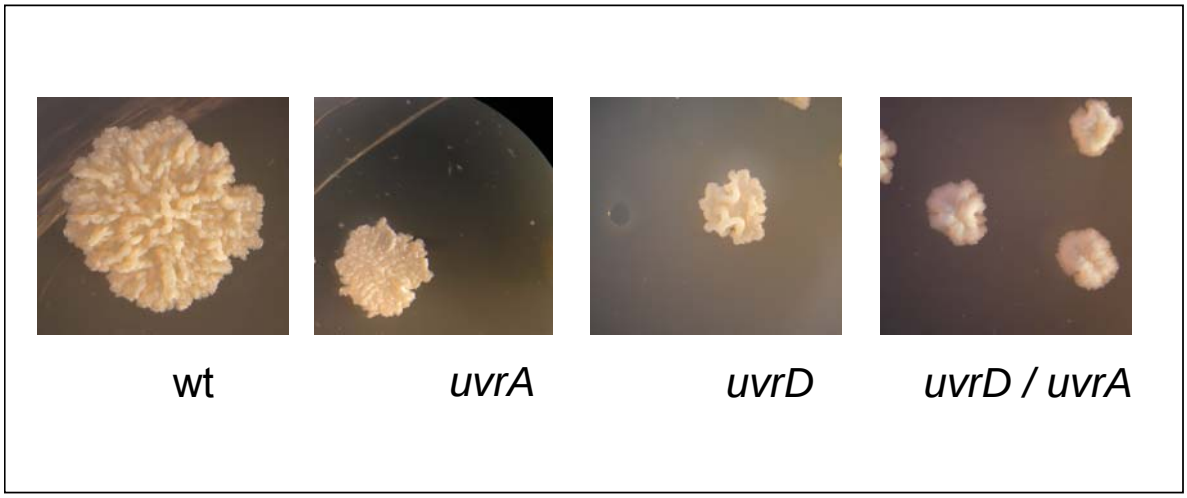


Figure 17: Colony morphology of *M. tuberculosis* wildtype (wt) and NER mutant strains

Colonies were grown on 7H10 agar for 40 days.

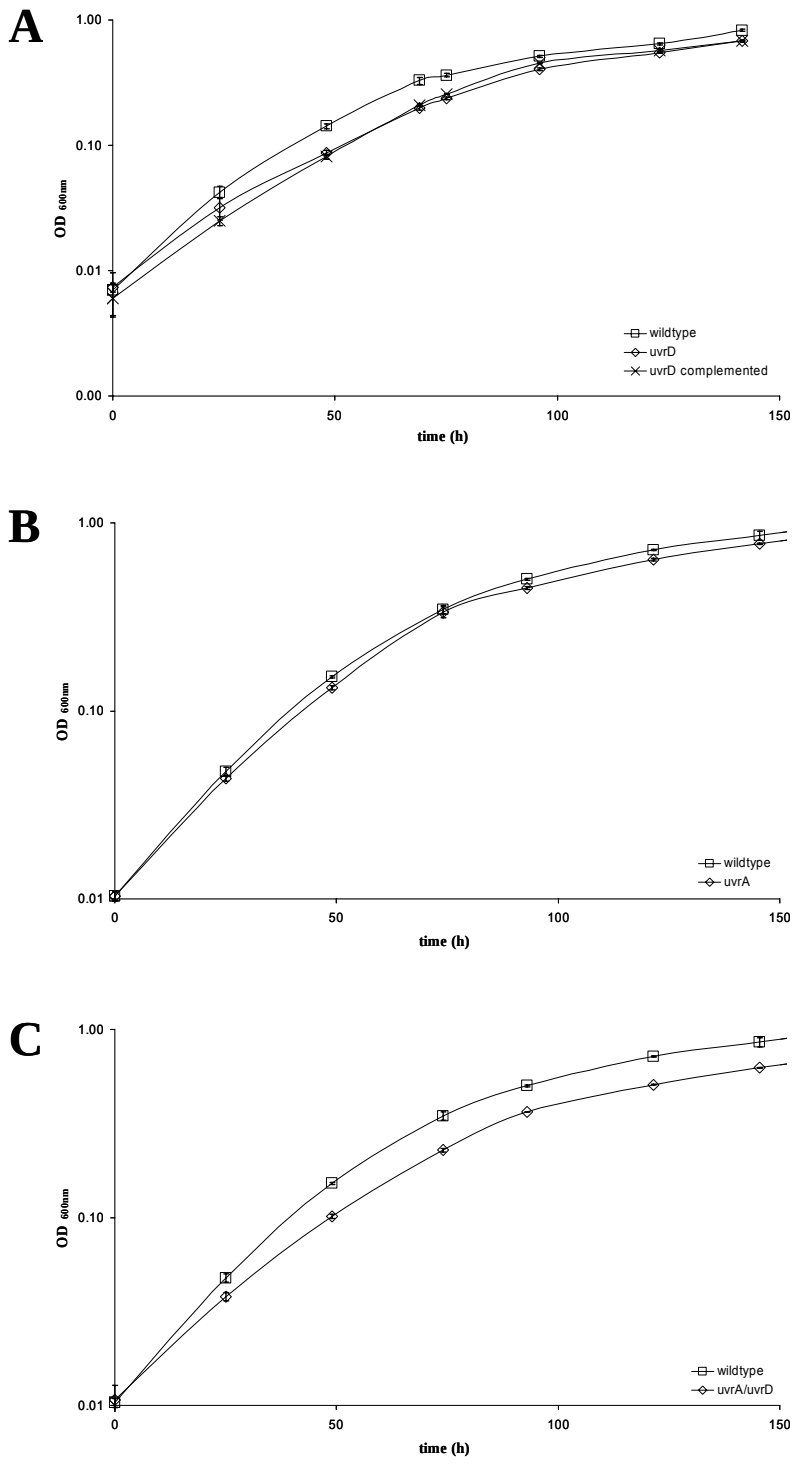


Figure 18: *In vitro* growth characteristics of *M. tuberculosis* NER mutants

Cultures were grown in 7H9 medium supplemented with OADC. (**A)** Growth of wildtype strain, *uvrD* mutant and complemented *uvrD* mutant **(B)** Growth of wildtype strain and *uvrA* mutant. **(C)** Growth of wildtype strain and *uvrA* / *uvrD* double mutant. Generation times: wildtype 22.6 h +/- 1.9, *uvrD* mutant 20.8 h +/- 0.2, complemented *uvrD* mutant: 22.2 h +/- 0.2, *uvrA* mutant: 25.7 h +/- 2.2, , *uvrA / uvrD* mutant: 36.9 h +/- 0.6.

4.5.2 Long-time survival of *M. tuberculosis uvrA / uvrD* mutant

In order to study survival of the *M. tuberculosis uvrA / uvrD* mutant in stationary phase, the *M. tuberculosis* wildtype and the *uvrA / uvrD* mutant were grown for 83 days and OD and CFU counts assessed at different time points.

Both strains reached viable-cell counts at around 1 x 10^8 CFU/ml after growth for nine days. After 83 days of incubation, the cell counts of the wildtype strain dropped to 3 x 10^7 CFU/ml. In contrast, the CFUs of the *uvrA / uvrD* mutant decreased to 2 x 10^5 CFU/ml, indicating an inability to survive during long-term stationary phase (Figure 19).

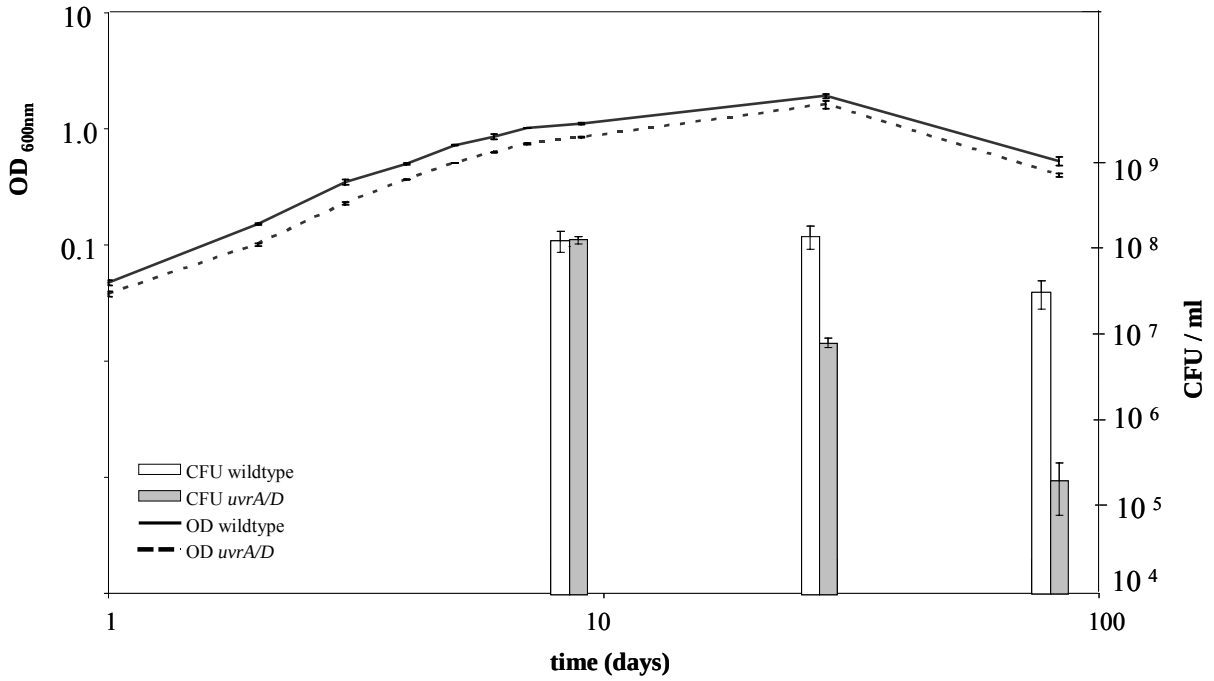


Figure 19: Long-term survival of *M. tuberculosis* wildtype and *uvrA / uvrD* mutant strain

OD values are plotted on the left axis, CFU counts are plotted on the right axis. All axes (including time on the x-axis) are plotted in logarithmic scale. Shown are the mean values of CFU counts of three independent cultures of each strain. Error bars: Standard deviation of the mean.

4.5.3 Susceptibility to DNA damaging agents

To test the ability of the mutant strains to cope with different DNA damaging agents, strains were challenged with DNA-damaging agents in a survival assay as detailed for the *M. smegmatis* NER mutants (section 4.3.3). *M. tuberculosis* NER mutant strains were assayed for their susceptibilities towards *tert*-butylhydroperoxide (TBH) and mitomycin C (MMC). In brief, three independent cultures were treated with the DNA-damaging agent for six days and CFU counts were determined by plating serial dilutions on 7H10 agar supplemented with OADC. Survival was calculated from the ratio of cells that were able to survive treatment compared to the untreated control.

TBH generates ROI and alkylating agents. The *uvrA* single and the *uvrA* / *uvrD* double mutant displayed a significant increase in susceptibility towards TBH compared to the wildtype strain. Susceptibility of the *uvrD* mutant strain was increased compared to the wildtype strain, although this was not found to be statistically significant (student's t-test, $p > 0.05$). Complementation of the *uvrD* mutant strain resulted in reconstitution of wildtype susceptibility (Figure 20 a).

Treatment with mitomycin C (MMC) results in inter- and intrastrand crosslinks. The *uvrA* single and the *uvrA* / *uvrD* double mutant displayed a significant increase in susceptibility towards MMC compared to the wildtype strain. In contrast, the *uvrD* mutant strain, the complemented *uvrD* mutant strain and the wildtype strains displayed similar susceptibilities (Figure 20 b, student's t-test, $p > 0.05$).

In general, these findings agree with those made in *M. smegmatis*. NER inactivation in both species results in similar phenotypes: compared to inactivation of *uvrB* or *uvrA*, deletion of *uvrD* had a much smaller effect on susceptibilities towards TBH and MMC. The strong effects of disruption of *M. smegmatis uvrB* and *M. tuberculosis uvrA* on TBH and MMC susceptibilities indicate that these proteins presumably have equivalent roles in both species.

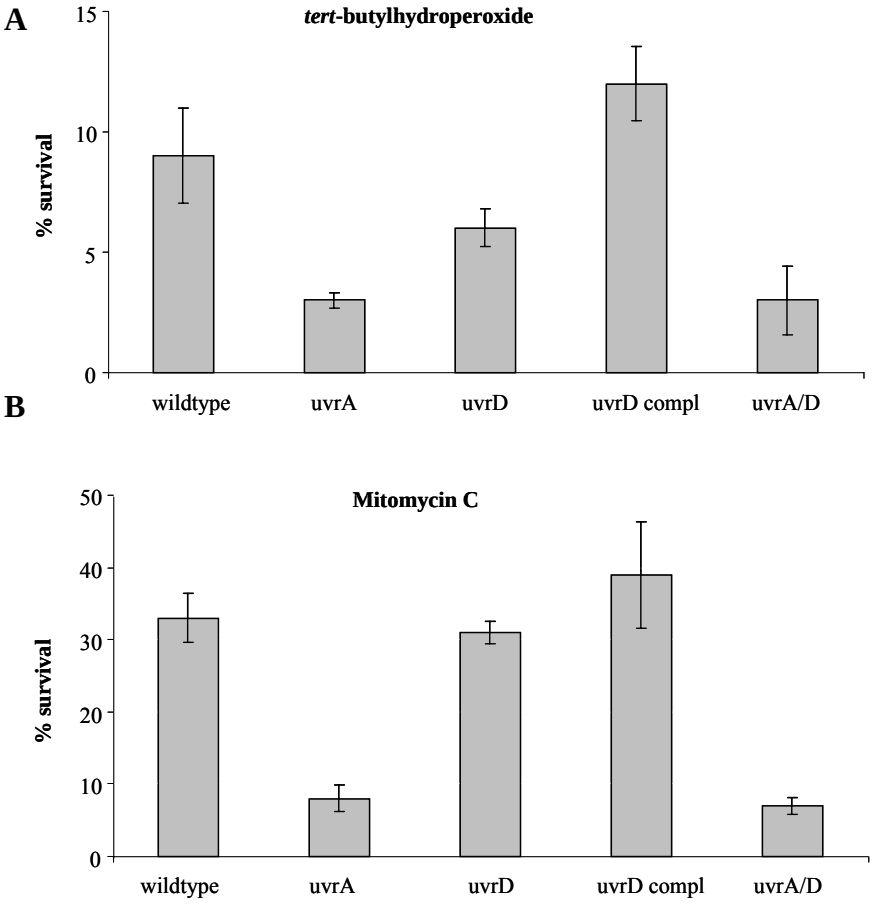


Figure 20: Susceptibility of *M. tuberculosis* NER mutants to DNA damaging agents

(A) *tert*-butyl hydroperoxide (250 µM) **(B)** mitomycin C (0.02 µg/ ml) Survival was calculated from the ratio of cells that survived treatment with the DNA damaging agent in proportion to CFU of the untreated control. Strains were incubated with the DNA damaging agent for 6 days. Shown are the mean values of one representative experiment performed with three independent cultures each. Standard deviations from the mean are given in the error bars. Abbreviations: *uvrA*, *uvrD*, *uvrA*/*uvrD* = *M. tuberculosis* NER mutant strains, *uvrD compl.* = *M. tuberculosis uvrD* strain complemented with the wildtype *uvrD* allele.

4.6 *In vivo* characterisation of *M. tuberculosis* NER mutant strains in a mouse model of infection

To evaluate the *in vivo* effect of NER disruption, the ability of the NER mutants to cause progressive infection in an *in vivo* mouse model was studied as described in section 3.12. In brief, groups of BALB/c mice were infected intravenously with approximately 5 x 10^5 cells of each strain. Progress of infection was assessed by determining organ CFU counts at various time points. Infection with wildtype *M. tuberculosis* cells resulted in progressive multiplication of bacteria in lungs and spleens, with the number of CFU reaching a plateau after nine weeks post infection. The *uvrA* mutant exhibited the same characteristic pattern of *in vivo* growth with organ CFU counts nearly identical to wildtype (Figure 21 a). In contrast, the *uvrD* mutant showed reduced ability to persist since the number of bacteria in lungs and spleens significantly decreased after nine weeks post infection (Figure 21 b).

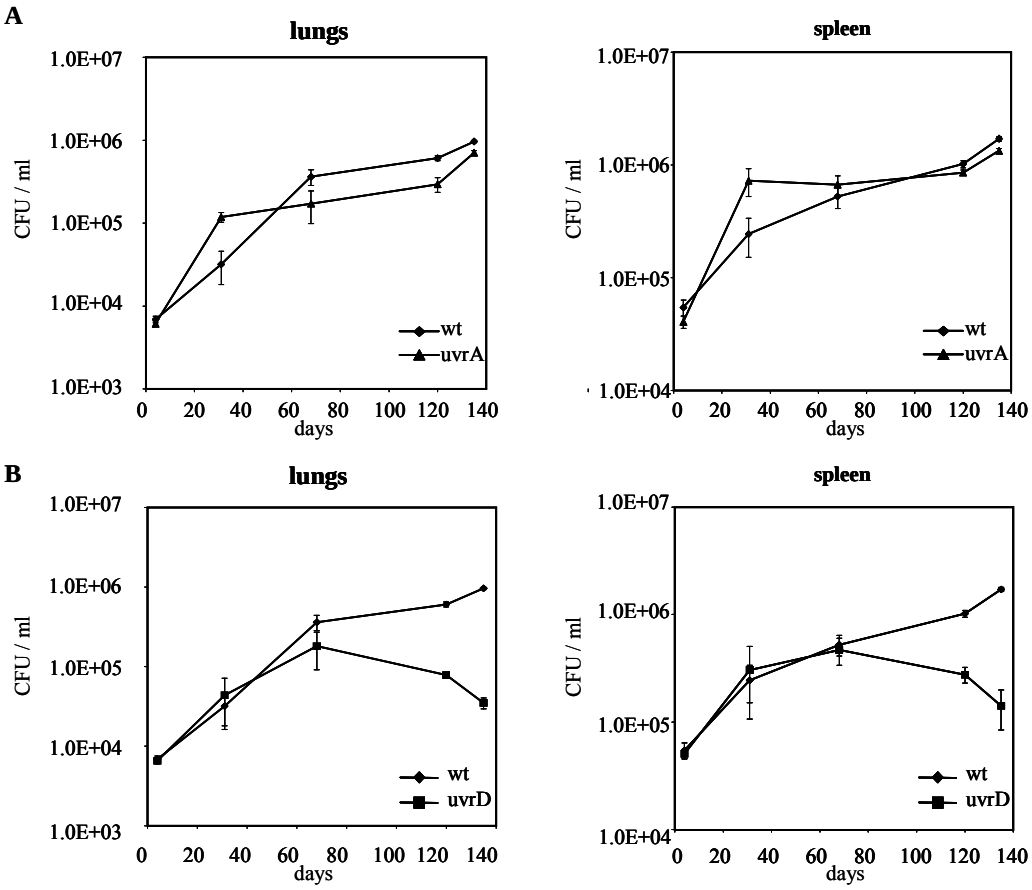


Figure 21: *In vivo* growth of *M. tuberculosis* wildtype strain, *uvrA* mutant strain (A) and *uvrD* mutant strain (B) in a mouse model of infection

Shown are CFU counts of lungs (left) and spleens (right) of intravenously infected BALB/c mice over a period of 140 days. CFU were calculated from the mean of CFU counts obtained from the lungs and spleens of three mice per strain for each time point. Error bars: standard deviations of the mean.

4.7 Generation of *M. tuberculosis* mutants with multiple deficiencies in DNA damage repair pathways

4.7.1 Failure to generate *uvrD2* knockout mutants

Genome analyses indicate that *M. tuberculosis* possesses a second DNA repair helicase assumed to be involved in NER. This homologue of *uvrD* has been annotated as *uvrD2*. To study the functional role of *uvrD2*, e.g. to analyse whether *uvrD2* is capable of compensating for the loss of *uvrD* function, we initiated the generation of a mutant deficient in *uvrD2*, as well as a mutant lacking both helicases. In parallel, we started the construction of a mutant deficient in *uvrA* and both helicases. Due to the limited number of selectable markers in *M. tuberculosis*, this strategy had to employ unmarked deletion mutagenesis.

M. tuberculosis wildtype, *uvrD* mutant strain and *uvrA* / *uvrD* double mutant strain were transformed with the *uvrD2* knockout vector. Transformants were isolated and amplified for genomic DNA preparation. Subsequently, DNA was subjected to Southern blot analysis making use of a probe specific for the upstream region of the *uvrD2* open reading frame. In the wildtype the probe hybridises to a 9 kbp fragment. Integration of the knockout vector occurs as a consequence of a single recombination event, resulting in a tandem arrangement of the functional *uvrD2* allele and the deletion allele. Vector integration can occur either upstream (5') or downstream (3') of the targeted gene and thus result in 5' single cross-over transformants or 3' single cross-over transformants, respectively. Expected fragments in 5' single cross-over transformants (5' sco) are 13.6 kbp and 3.9 kbp; in 3' single cross-over transformants (3' sco) 8.1 kpb and 9.4 kbp. Successful disruption of *uvrD2* would result in a hybridization signal at 7.5 kbp (a representative Southern blot is shown in Figure 22).

In *M. tuberculosis* wildtype transformants, we obtained 4 clones that displayed a pattern indicative of a single cross-over event (all 3' sco). Analyses of clones derived from transformation of the *M. tuberculosis uvrD* mutant strain revealed that 4 clones had undergone a single recombination event (all 3' sco). Transformation of the *M. tuberculosis uvrA / uvrD* mutant yielded two single cross-over transformants (1 x 5'sco and 1 x 3' sco). To inactivate *uvrD2*, a second intramolecular recombination event has to occur in the single cross-over transformant. Single cross-over clones were grown for ten days in liquid medium to allow for homologous recombination. Subsequently, the bacteria were plated on selective 7H10 agar containing streptomycin. This counterselection selects against maintenance of knockout vector sequences which contain the wildtype *rpsL* gene. Loss of this allele upon

intramolecular recombination renders cells resistant to streptomycin. In order to calculate recombination frequencies, serial dilutions were plated on permissive medium without streptomycin. Frequencies for vector integration by intermolecular single cross-over recombination in mycobacteria are in the range of 10^{-4} to 10^{-6} (Sander *et al.*, 1995). Expected frequencies for the second intramolecular recombination are in the range of 10^{-3} to 10^{-4}. Counterselections resulted in the expected ratios of streptomycin-resistant clones, as summarised in Table 6. Despite the presence of a high number of streptomycin-resistant clones, Southern blot analyses demonstrated that no second intramolecular recombination event has occurred in any clone analysed. As all clones displayed a hybridization signal pattern indicative of tandem arrangement of the functional *uvrD2* allele and the deletion allele (i.e. single cross-over clones), the resistance to streptomycin was due to spontaneous point mutations in the *rpsL* gene or due to gene conversion of the vector encoded susceptible *rpsL* allele by the chromosomal resistant *rpsL* allele. In total, 144 clones were analysed for inactivation of *uvrD2* in the wildtype strain, 72 clones were analysed for inactivation of *uvrD2* in the *uvrD* mutant strain, 24 clones were analysed for *uvrD2* loss in the *uvrA* / *uvrD* mutant strain. In none, an *uvrD2* knockout mutant was obtained.

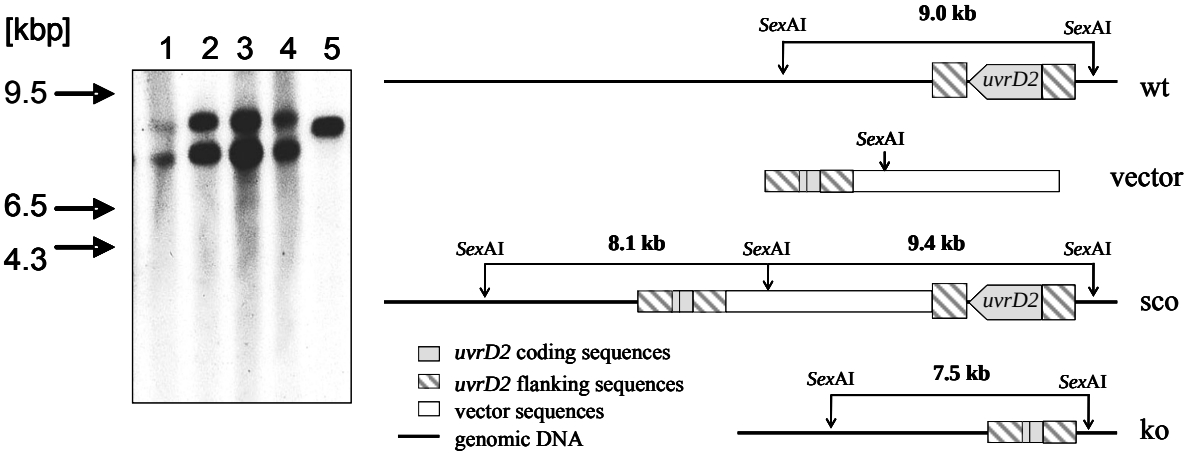


Figure 22: Failure to generate *uvrD2* mutants in *M. tuberculosis*

Left panel: Southern blot analysis. Genomic DNA of clones obtained after counterselection of *uvrD2* single cross-over mutants (lane 1-4) and DNA of the wildtype strain (lane 5) were digested with *Sex*AI and hybridised with a probe specific for the upstream region of *uvrD2*. The hybridization signals in lane 1-4 reveal a pattern indicative for single cross-over clones, i.e. no *uvrD2* knockout mutant could be obtained after counterselection.

Right panel: Schematic illustration of the *uvrD2* locus and the Southern blot strategy. Shown are the genomic organisation of wildtype (wt), the knockout vector that contains the deletion allele (vector), the single cross-over genotype (sco) and the mutated genomic *uvrD2* region in the knockout mutant (ko). Fragments detected by the probe specific for the 5' flanking region are shown in bold letters.

Table 6: Ratios of streptomycin-resistant clones after counterselection for *uvrD2* inactivation

Parental strain	Ratio of streptomycin resistant clones		Tested clones / obtained knockouts
wildtype	1st counterselection	2.1 x 10^{-4}	48 / 0
	2nd counterselection	1.4 x 10^{-4}	48 / 0
	3rd counterselection	3.1 x 10^{-4}	48 / 0
uvrD	1st counterselection	3.6 x 10^{-4}	36 / 0
	2nd counterselection	2.5 x 10^{-3}	36 / 0
uvrA / uvrD	1st counterselection	1.8 x 10^{-3}	24 / 0

4.7.2 Failure to generate *M. tuberculosis* NER mutants deficient in *recA*

To study the impact of deficiency in more than one DNA repair pathway, we aimed to construct *M. tuberculosis* mutants that are devoid of the NER and homologous recombination repair pathway. Therefore, we transformed the *recA* knockout vector into the *M. tuberculosis uvrD* mutant strain and into the *uvrA / uvrD* mutant strain, respectively. *RecA* single cross-over transformants were obtained and subjected to counterselection.

Overall, 3 independent counterselections were performed, in total 102 clones were screened (table 7). Counterselection of *recA* single cross-over transformants did not result in successful disruption of *recA*.

Table 7: Ratios of streptomycin-resistant clones after counterselection for *recA* inactivation

Parental strain	Ratio of streptomycin resistant clones		Tested clones / obtained knockouts
uvrD	1st counterselection	1.2 x 10^{-3}	30 / 0
	2nd counterselection	4.3 x 10^{-5}	36 / 0
	3rd counterselection	3.2 x 10^{-3}	36 / 0
uvrA / uvrD	No sco obtained after transformation (22 clones tested)		

4.8 SMC in mycobacteria

Analyses of molecular databases indicate that all mycobacterial genomes contain an SMC homologue. It is even present in *M. leprae*, a bacterium that has lost various gene functions during reductive genomic evolution (Cole *et al.*, 2001). Mycobacterial SMC proteins share all those structural motifs conserved in members of the SMC family as determined by *in silico* analyses (Clustal W 1.83 multiple sequence alignment and InterPro Scan protein motive search). Homologous sequences are restricted to the head, hinge and tail domains. The three domains are connected by two long heptad repeat regions predicted to form coiled-coils (Figure 23).

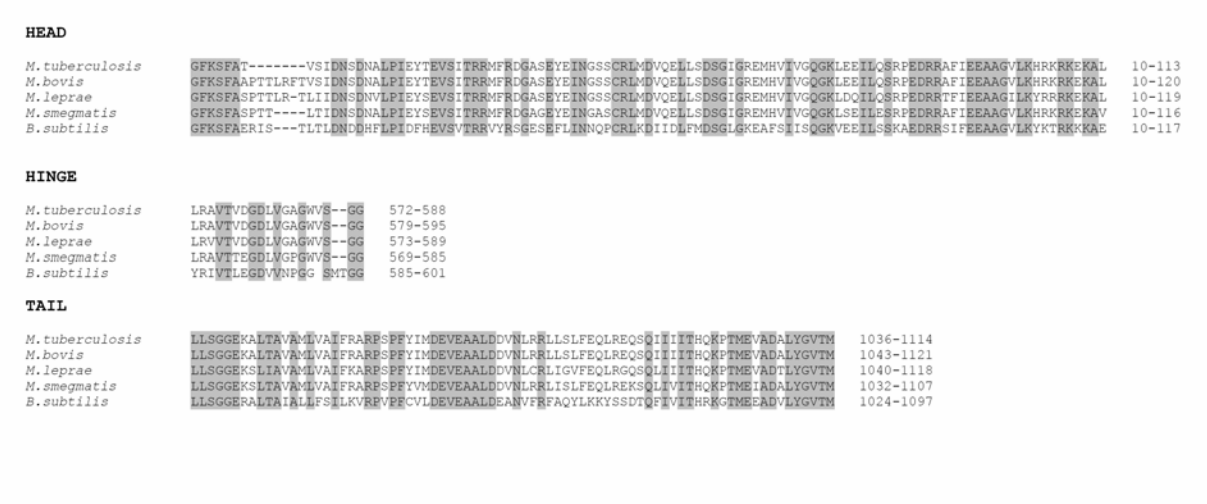


Figure 23: Sequence alignments of the three conserved SMC regions

The conceptual *M. smegmatis* SMC ORF encodes an 1195 amino-acid polypeptide with a calculated molecular mass of 129.7 kDa. Overall, *M. smegmatis* SMC shares 74% identity with the 1205 amino-acid *M. tuberculosis* SMC and 55 % identity with the 1186 amino-acid SMC sequence of *B. subtilis*. The SMC protein of *M. tuberculosis* has a calculated molecular mass of 130.6 kDa.

Eukaryotic SMC proteins functionally interact with other proteins, and SMC of *B. subtilis* was found to function in concert with two proteins called ScpA and ScpB (Hirano and Hirano, 2004). Both, *M. tuberculosis* and *M. smegmatis*, have homologues of these interacting proteins, suggesting that mycobacterial SMC is functional (ScpA homologues: Rv1709/MSMEG3748; ScpB homologues: Rv1710/MSMEG3749, data not shown). All sequences were obtained from the Institute for Genomic research (TIGR) website and TubercuList of Institute Pasteur respectively.

4.9 Generation of *M. smegmatis smc* mutants

The *M. smegmatis smc* knockout mutant was constructed by allelic exchange mutagenesis making use of the dominant-negative selectable marker *rpsL* as described in section 3.10 (Sander *et al.*, 1995). Using genomic DNA as template, a 2 kb deletion allele was constructed by joining PCR products that contain parts of the open reading frame flanked by upstream and downstream sequences. The resulting deletion allele lacks basepairs 747 to 2660 of the 3588 basepair open reading frame. To allow proper selection of gene deletion mutants, a hygromycin resistance cassette was inserted between the two fragments resulting in the deletion allele *smc::hyg*. One prominent phenotype of previously described *smc* null mutants is temperature sensitivity (Britton *et al.*, 1998; Jensen and Shapiro, 1999; Volkov *et al.*, 2003). Therefore, transformation and counterselection procedures of *M. smegmatis* were performed at room temperature. All selection steps were set off with bacteria taken from frozen stocks to ensure that no suppressor mutations arise due to extensive passaging. Disruption of the *smc* gene was verified by Southern blot analysis (Figure 24). This analysis revealed a 3.8 kbp hybridization signal in the mutant strain compared to a 7.3 kbp fragment in the parental strain. The shift in fragment size in the *smc* mutant strain results from replacement of a 1914 bp *smc* gene fragment by a 1758 bp hygromycin cassette, which provides a further restriction site.

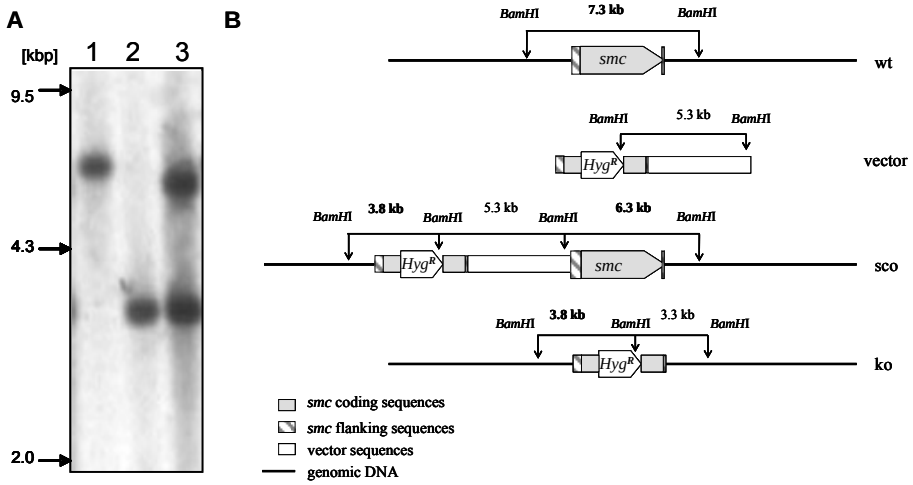


Figure 24: Generation of *M. smegmatis smc* mutant

(A) Southern blot analysis. Genomic DNA from *M. smegmatis* wildtype (lane 1), *smc* mutant (lane 2) and the *smc* single cross-over mutant (lane 3) was digested with *BamH*I and probed with a 685 bp *Pvu*II gene fragment containing the 5' flanking region of the *smc* gene. The presence of a single 3.8 kbp fragment in the mutant strain instead of a 7.3 kbp fragment observed in the parental strain demonstrates successful deletion of *smc* coding sequences.

(B) Schematic illustration of the smc locus and the Southern blot strategy. Shown are the genomic organisation of wildtype (wt), the knockout vector that contains the deletion allele *smc::hyg* (vector), the single cross-over genotype (sco) and the mutated genomic *smc* region in the knockout mutant (ko). Fragments detected by the probe specific for the 5' flanking region are shown in bold letters.

4.10 Characterisation of *M. smegmatis smc* mutant strain

4.10.1 *In vitro* growth

In order to analyse the phenotype of the *smc* mutant strain, the *in vitro* growth characteristics were determined in 7H9 broth and compared to the parental wildtype strain.

One main feature of all *smc* null mutants that have been previously described is temperature sensitive growth. Therefore, besides characterising growth at standard incubation temperature (37°C), the growth rates at lower (30°C) and higher temperatures (42°C) were also assessed.

In contrast to *smc* null mutant phenotypes reported for *B. subtilis* and *C. crescentus*, the growth rates of the mutant and the wildtype were similar in *M. smegmatis* (Figure 25). Doubling times in logarithmic phase were nearly identical indicating that *smc* deletion does not affect *in vitro* growth of *M. smegmatis*. Additionally, wildtype and mutant were indistinguishable with respect to colony morphology; when grown on 7H10 agar, both strains showed the typical features of mycobacterial colonies with rough dry surfaces and irregular edges (data not shown).

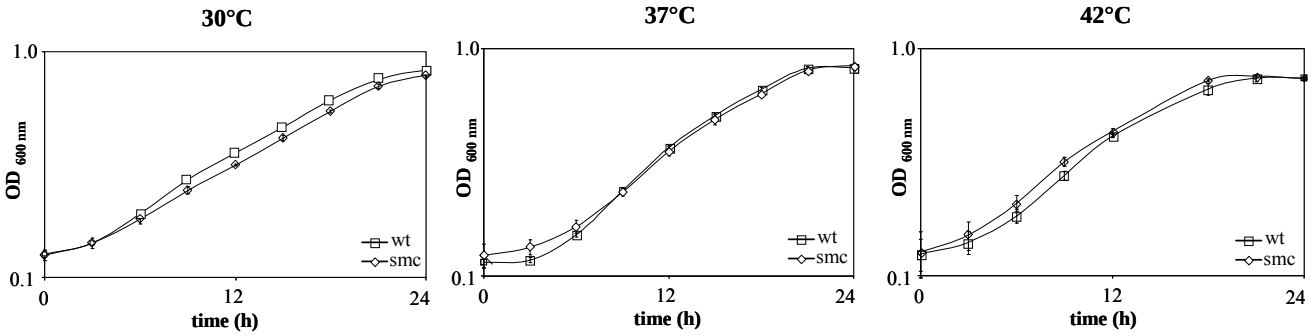


Figure 25: Growth of *M. smegmatis* wildtype strain and *M. smegmatis smc* mutant strain at various growth temperatures

Doubling times: 30°C: wildtype 4.0 h +/- 0.3, *smc* mutant 4.0 h +/- 0.2; 37°C: wildtype 3.6 h +/- 0.05, *smc* mutant 3.5 h +/- 0.1; 42°C: wildtype 5.0 h +/- 0.6, *smc* mutant 5.4 h +/- 0.5.

4.10.2 Spontaneous mutation frequencies

To assess whether mutation in SMC leads to an increase of the spontaneous mutation frequency, the frequencies of mutation to rifampicin resistance were determined as described in section 3.11.1.

Mutation frequencies were comparable in wildtype ($3.1 \times 10^{-8} \pm 1.5 \times 10^{-8}$) and mutant strain ($3.8 \times 10^{-8} \pm 1.0 \times 10^{-8}$; $p = 0.68$; $n = 4$; student's t-test), indicating that SMC deficiency does not result in a mutator phenotype in *M. smegmatis*.

4.10.3 Susceptibility to ofloxacin, a DNA gyrase inhibitor

DNA gyrase, which generates negative supercoiling, and topoisomerase I, which prevents excessive negative supercoiling by DNA gyrase, are essential to maintain the optimal topological state of DNA in the cell (Champoux, 2001). Previous investigations indicate that SMC proteins mediate chromosome compaction by contributing to the introduction of negative supercoils into DNA (Tadesse and Graumann, 2006). This hypothesis is corroborated by the observation that SMC and MukB mutants are hypersusceptible to gyrase inhibitors (Sawitzke and Austin, 2000), while loss of MukB function can be partially compensated by reducing topoisomerase I activity (Tadesse and Graumann, 2006).

To investigate if SMC also interferes with DNA supercoiling in *M. smegmatis*, the minimal inhibitory concentration (MIC) towards the gyrase inhibitor ofloxacin using E-test (AB BIODISK) was determined. In contrast to findings made in other organisms (Sawitzke and Austin, 2000), *M. smegmatis* wildtype and *smc* mutant displayed identical drug susceptibility to ofloxacin with a MIC of 0.19 μg/ml.

4.10.4 Survival after treatment with mitomycin C

To investigate whether the *M. smegmatis smc* mutant is impaired in DNA repair, susceptibility to mitomycin C was studied. This DNA damaging agent generates intrastrand crosslinks. No difference in susceptibility of the wildtype and the *smc* mutant strain was observed, implying that mycobacterial SMC has a minor role in DNA repair (data not shown).

4.10.5 Susceptibility to ionizing radiation

Since SMC in other organisms is known to participate in the repair of double-strand breaks (Potts *et al.*, 2006; Sjogren and Nasmyth, 2001), the ability of the SMC mutant strain to survive treatment with ionizing radiation was assessed. Midlog cultures were irradiated and CFU determined as described in section 3.11.5. We also included a *recA* deficient mutant strain in this experiment. RecA is involved in the repair of double-strand breaks by homologous recombination.

In contrast to the *recA* mutant strain, no major difference in survival between wildtype strain and *smc* mutant strain was found (Figure 26).

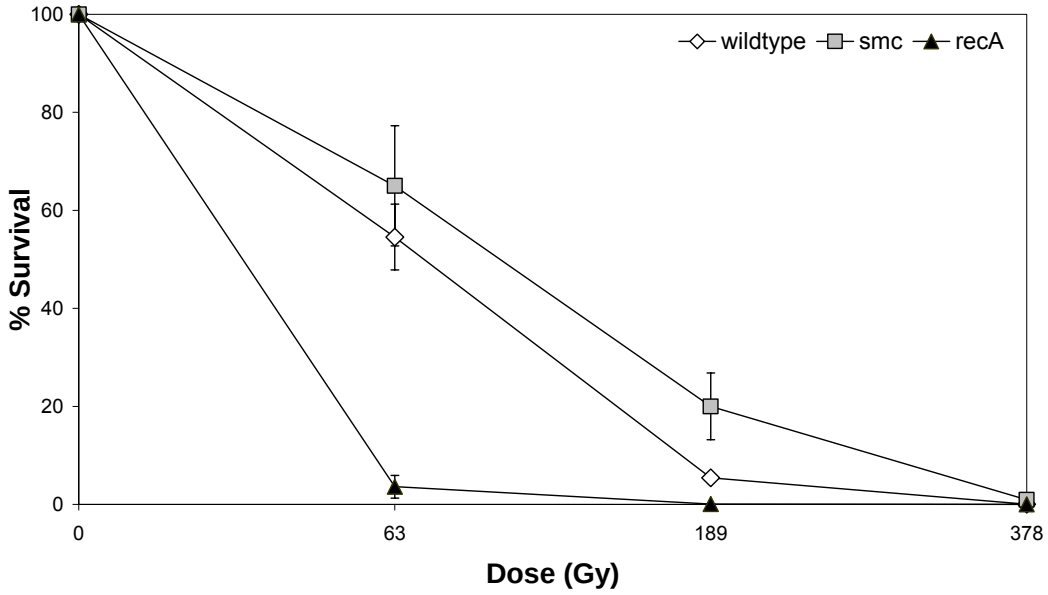


Figure 26: Survival of *M. smegmatis* wildtype, *smc* and *recA* mutant strains following exposure to ionizing radiation

Survival was calculated from the ratio of cells that survived treatment with ionizing radiation in proportion to the untreated control. Shown are the mean values of one representative experiment performed with three independent cultures each. Standard deviations of the means are given in the error bars.

4.10.6 Susceptibility to UV-C

The eukaryotic SMC5/6 complex has been described to act in a second pathway for the postreplicative repair of UV-induced lesions, which is different from the standard nucleotide excision repair (NER) pathway (Lehmann *et al.*, 1995). In order to analyse the function of mycobacterial SMC in the repair of UV-induced lesions, sensitivity of the *smc* mutant strain to UV light was studied. Three independent cultures of each strain were grown in 7H9 medium until late log phase (OD 1-2). Serial dilutions of each strain were plated on LB agar. Subsequently, the plates were exposed to different UV doses (0 - 160 mJ/cm^2, 254 nm) and the number of CFUs were determined. In contrast to the *recA* mutant strain which was severely impaired in survival, the wildtype and the *smc* mutant strain displayed similar susceptibilities to UV (Figure 27).

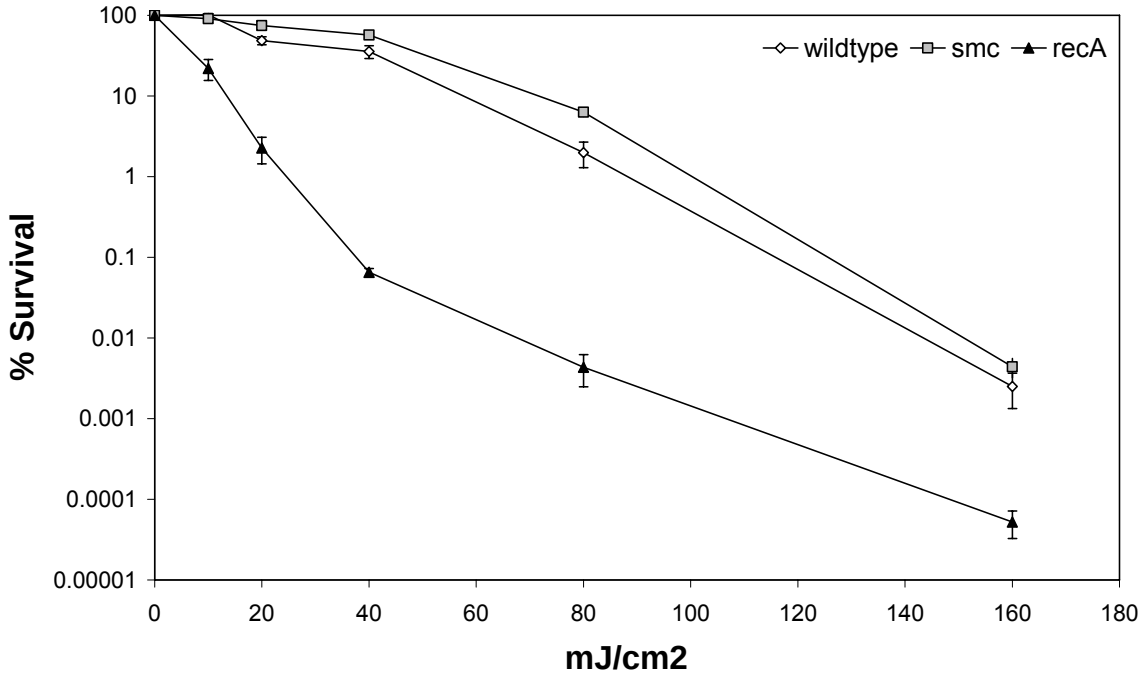


Figure 27: Survival of *M. smegmatis* wildtype, *smc* and *recA* mutant strains following exposure to UV

Survival was calculated from the ratio of cells that survived treatment with UV-C in proportion to the untreated control. Shown are the mean values of one representative experiment performed with three independent cultures each. Standard deviations of the means are given in the error bars. Ratio of treated to untreated cells is plotted in logarithmic scale.

4.10.7 Survival in stationary phase

Some genes are not essential for exponential growth but a prerequisite for remaining viable during prolonged periods of oxygen and nutritional deprivation. In order to gain insight into a possible role of SMC for survival of *M. smegmatis* in stationary phase, CFU counts of the *M. smegmatis* wildtype and *smc* mutant grown for 28 days in 7H9 broth were compared.
Both strains reached viable cell counts in the vicinity of 1 x 10^9 CFU/ml after growth for two days. These counts dropped to approximately 5 x 10^8 CFU/ml upon further incubation and remained virtually unchanged for the next 28 days (Figure 28).

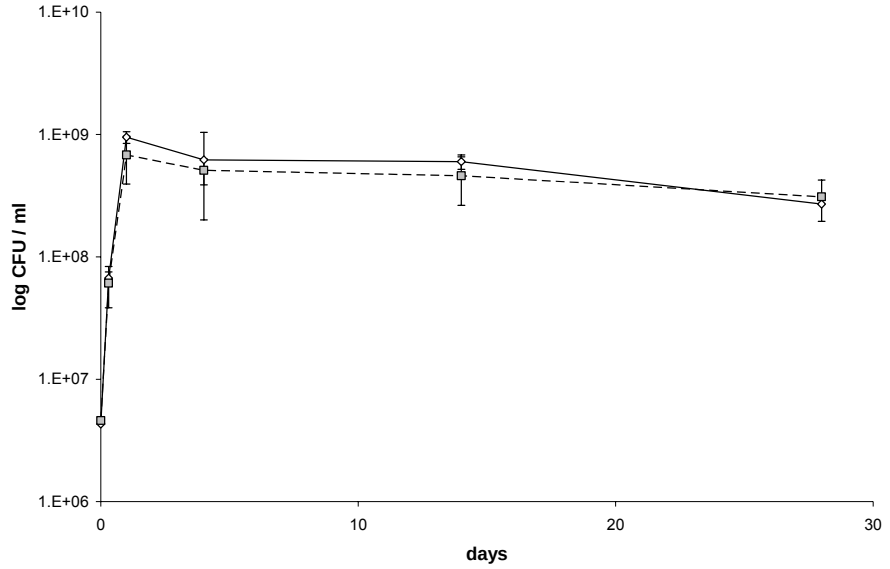


Figure 28: Long-term survival of *M. smegmatis* wildtype (solid lines) and *smc* mutant (dashed lines) strain.
Shown are the mean values of CFU counts of three independent cultures of each strain.

4.11 Generation of *M. tuberculosis smc* mutants

The *M. tuberculosis smc* knockout mutant was constructed by allelic exchange mutagenesis as described in section 3.10. A deletion allele lacking 2121 basepairs of the 3618 basepair *smc* open reading frame (the resulting mutant lacks basepairs 553-2674 of the ORF) was constructed by allelic replacement mutagenesis as outlined for *M. smegmatis* in section 4.9. Disruption of the *smc* gene was controlled by Southern blot analysis (Figure 29). This analysis revealed a 4.3 kbp hybridization signal in the mutant strain compared to an 11.1 kbp fragment in the parental strain. The difference in size is due to deletion of the genomic *smc* region and introduction of an additional *BamH*I restriction site provided in the hygromycin resistance cassette.

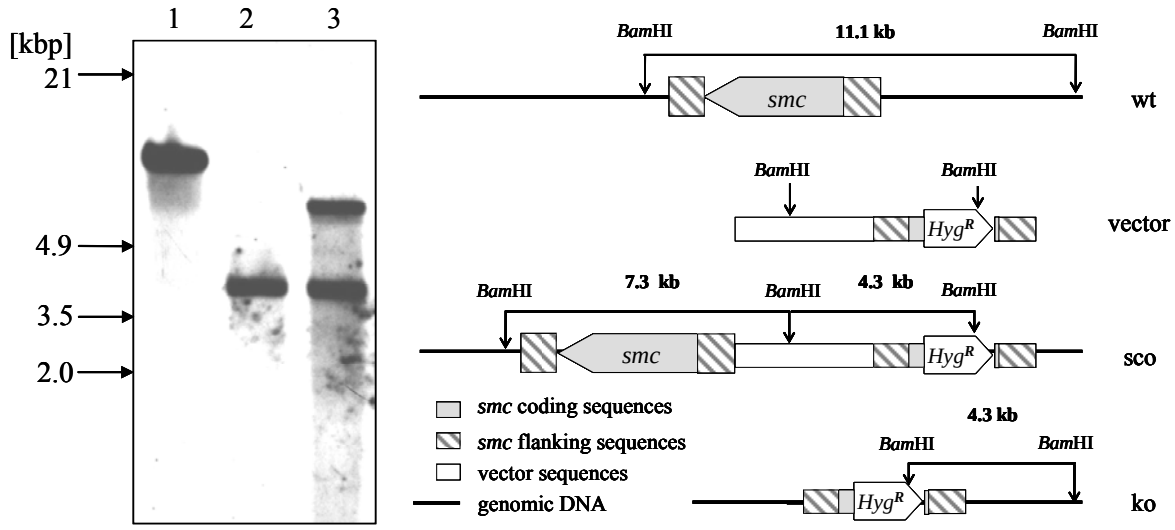


Figure 29: Generation of *M. tuberculosis smc* mutant

Left panel: Genomic DNA from *M. tuberculosis* wildtype (lane 1), *M. tuberculosis smc* mutant (lane 2) and *M. tuberculosis smc* sco (lane 3) was digested with *BamH*I and probed with a 935 bp *Nde*I DNA fragment containing the 5' flanking regions of the *smc* gene. The presence of a single 4.3 kbp fragment instead of a 11.1 kbp fragment as seen in the parental strain demonstrates successful deletion of *smc* coding sequences.
Right panel: Schematic illustration of the smc locus and the Southern blot strategy. Shown are the genomic organisation of wildtype (wt), the knockout vector that contains the deletion allele *smc::hyg* (vector), the single cross-over genotype (sco) and the mutated genomic *smc* region in the knockout mutant (ko). Fragments detected by the probe specific for the 5' flanking region are shown in bold letters.

4.12 Characterisation of *M. tuberculosis smc* mutant strain

4.12.1 *In vitro* growth

In vitro growth characteristics of the *M. tuberculosis smc* mutant strain were analysed in 7H9 broth supplemented with OADC at 37°C and compared to the parental wildtype strain. No difference in growth between the mutant and the wildtype was found. Doubling times in logarithmic phase were nearly identical indicating that interruption of *smc* does not affect *in vitro* growth of *M. tuberculosis* (Figure 30 a). Additionally, wildtype and mutant were indistinguishable with respect to colony morphology; when grown on 7H10 agar, both strains showed the typical features of mycobacterial colonies with rough dry surfaces and irregular edges (Figure 30 b).

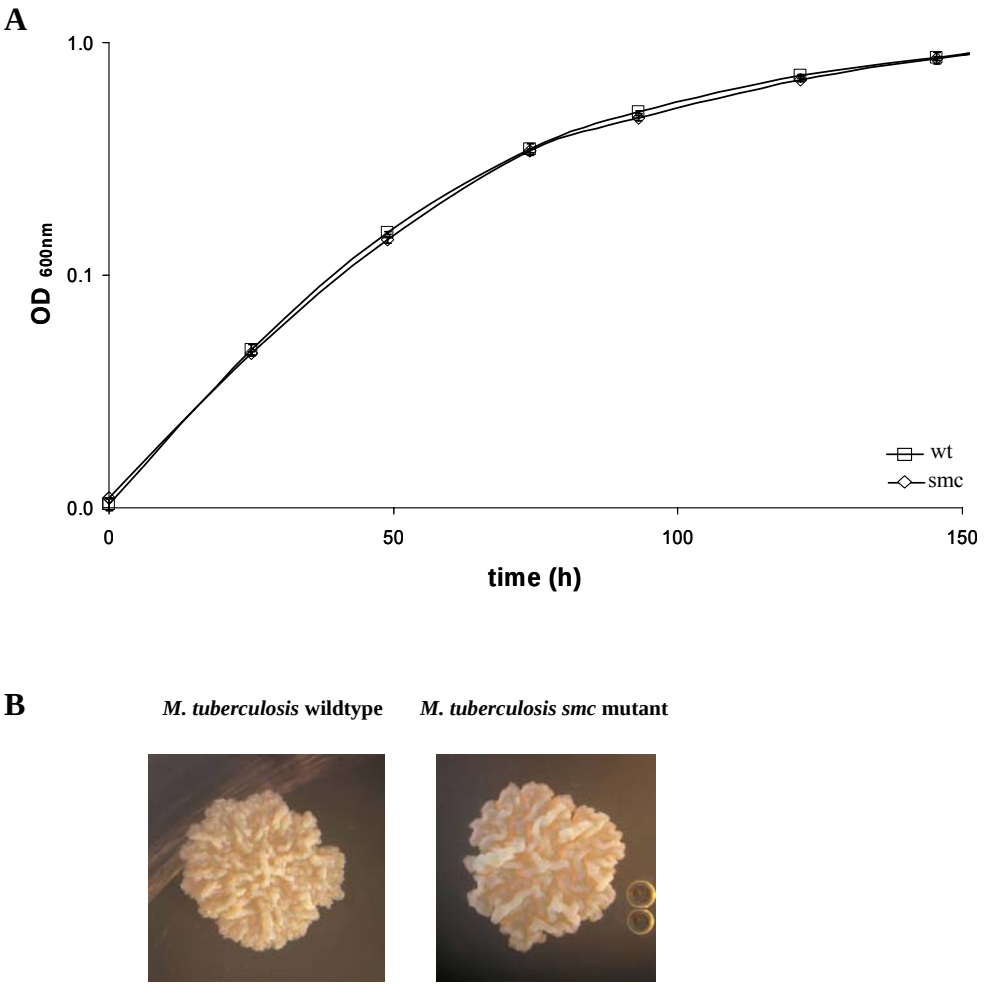


Figure 30: Growth analysis of M. *tuberculosis* wildtype and *smc* mutant strain

A Growth curve. Doubling times: wildtype 22.6 h +/- 0.9 h, *smc* mutant strain 23.4 h +/- 0.8.

B Colony morphology. Colonies were grown on 7H10 agar for 40 days.

4.12.2 Survival in stationary phase

In order to gain insight into a possible role of SMC for survival of *M. tuberculosis* in stationary phase, CFU counts of the *M. tuberculosis* wildtype and *smc* mutant grown for 90 days in 7H9 broth supplemented with OADC were compared.

Both strains reached viable-cell counts in the vicinity of 1 x 10^8 CFU/ml after growth for nine days. These counts remained virtually unchanged for the next 20 days. After 83 days of incubation, the CFUs of both strains dropped to approximately 3 x 10^7 CFU/ml (Figure 31).

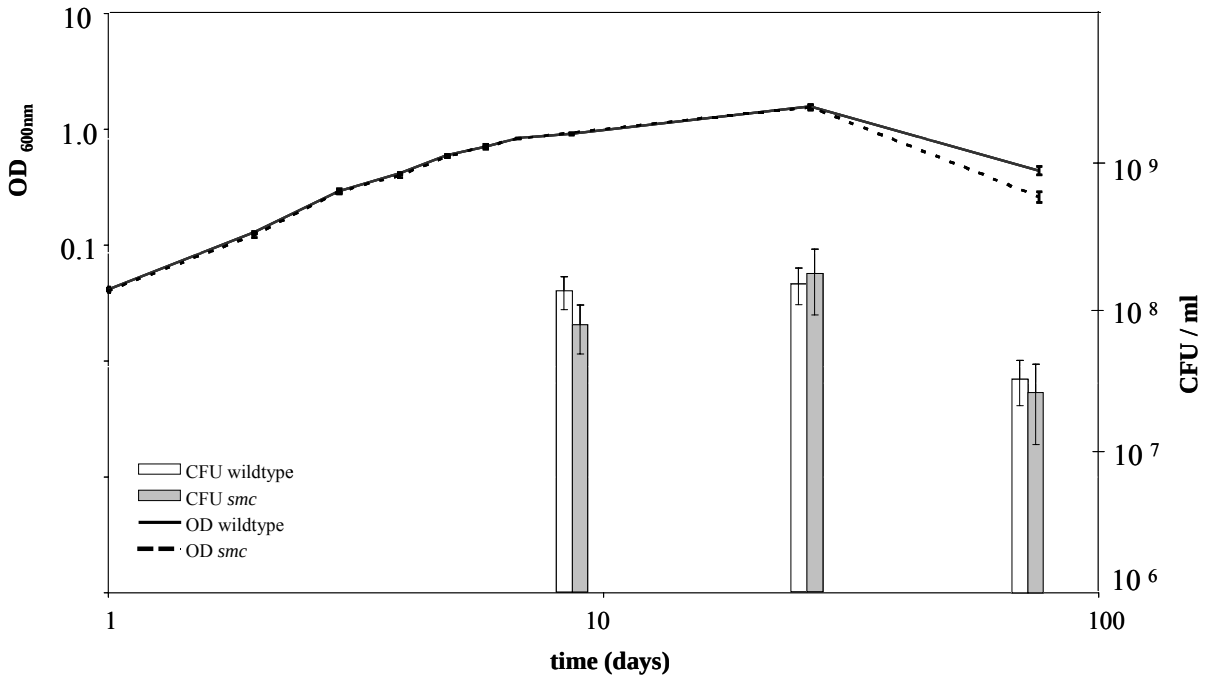


Figure 31: Long-term survival of *M. tuberculosis* wildtype and *smc* mutant strain

OD values are plotted on the left axis, CFU counts are plotted on the right axis. All axes (including time on the x-axis) are plotted in logarithmic scale. Shown are the mean values of three independent cultures of each strain. Error bars: standard deviation of the mean.

5 Discussion

5.1 DNA repair in mycobacteria

Annually, *Mycobacterium tuberculosis* is responsible for nearly two million deaths (data from "Tuberculosis facts- Handout 2007" published on www.who.int). Part of its success lies in its ability to survive and replicate in macrophages, but the molecular mechanisms allowing it to do so are not yet understood (Kaufmann, 2001).

The production of reactive oxygen intermediates and reactive nitrogen intermediates by the macrophages, once activated, is regarded an effective defence mechanism against intracellular pathogens (Chan *et al.*, 1992; Ding *et al.*, 1988; Nathan and Shiloh, 2000). However, *M. tuberculosis* seems to be superbly adapted to these extremely unfavourable conditions which demand efficient mechanisms to ensure genome stability. Despite the necessity for repair of DNA damage, mycobacteria are devoid of the *mutLS*-based post-replicative mismatch repair (MMR) system (Mizrahi and Andersen, 1998). MMR systems are highly conserved throughout evolution and contribute to mutation avoidance by correcting replication errors resulting from nucleotide misincorporation and polymerase slippage (Kunkel and Erie, 2005; Schofield and Hsieh, 2003). In addition, MMR inhibits recombination between non-identical (homeologous) sequences and thus takes part in ensuring the fidelity of recombination (Matic *et al.*, 1995; Rayssiguier *et al.*, 1989; Worth *et al.*, 1994). The only MMR homologue present in mycobacteria is a DNA-helicase named UvrD, which is also involved in NER.

Previous studies provided evidence that absence of MMR in mycobacteria does not affect genome fidelity as they display spontaneous mutation rates comparable with those of MMR proficient species (Springer *et al.*, 2004). Assuming that alternative DNA repair mechanisms compensate for the lack of MMR, the role of NER in mycobacteria was studied. Prokaryotic NER has been extensively analysed in *E. coli* (Sancar, 1996). This DNA repair pathway is accomplished by the products of the *uvrABC* and *uvrD* genes and is capable of repairing a wide range of damage types (Sancar, 1996). Analysis of NER in *Bacillus subtilis* (Sancar, 1996), *Streptococcus pneumoniae* (Sicard *et al.*, 1992), *Mycoplasma genitalium* (Sancar, 1996) and *Deinococcus radiodurans* (Minton, 1994) suggest that the repair mechanism originally characterised in *E. coli* is highly conserved in prokaryotes. However, little is known about nucleotide excision repair in mycobacteria. *In silico* genome analyses provided evidence that mycobacteria have a full set of NER genes including *uvrABC* and *uvrD* (Mizrahi and Andersen, 1998).

5.2 *M. smegmatis* NER mutants

To analyse the role of the NER system in mycobacteria, we constructed *M. smegmatis* mutants deficient in the NER components *uvrD* and *uvrB*, as well as a mutant deficient in both proteins. Analysis of susceptibilities towards DNA damaging agents revealed that disruption of NER in *M. smegmatis* results in an impairment of DNA repair. NER mutants in *E. coli* are sensitive not only to irradiation with UV, but also to a wide range of chemical agents, including mitomycin C (Friedberg, 2005). Consistently, we found that mycobacterial NER mutants are sensitive to treatment with UV and mitomycin C. The *M. smegmatis uvrB* and *uvrB /uvrD* double mutant, but not the *uvrD* single mutant are extremely susceptible towards MMC, corroborating previous observations that *E. coli uvrB* mutants were significantly more susceptible towards MMC than other *E. coli uvr* mutants (Vidal *et al.*, 2006). The *M. smegmatis* NER mutants, but not the *recA* mutant, are hypersensitive to TBH, indicating a major role of mycobacterial NER in the repair of oxidative and alkylation damage. In contrast, repair of oxidative DNA damage is primarily accomplished by base excision repair (BER) and homologous recombinational repair in *E. coli* (Konola *et al.*, 2000). NER is not involved in the repair of ROI induced lesions here, as *E. coli uvrA, uvrB* and *uvrC* mutants display wildtype susceptibility to hydroperoxide (Imlay and Linn, 1987); alkylation damage is mainly repaired by methyltransferases and BER in *E. coli* (Nowosielska et al., 2006) - NER is, if at all, only moderately capable of repairing these lesions (Samson *et al.*, 1988; Van Houten *et al.*, 2005). Our findings indicate that in contrast to the observations made in *E. coli*, NER is more important than homologous recombinational repair in defence against ROI and alkylation damage in mycobacteria.
Repair of RNI induced DNA damage in *E. coli* is mainly accomplished by repair mechanisms other than NER, primarily by BER and homologous recombinational repair (Spek *et al.*, 2001), although a role for UvrA in resistance to acidified sodium nitrite has been reported (Sidorkina *et al.*, 1997). The mycobacterial NER mutants are more susceptible than both the wildtype strain and the *recA* mutant to the RNI generating agent acidified sodium nitrite. The mild phenotype RNI/ROI phenotype of the *recA* mutant indicates a minor role for homologous recombinational repair in repairing ROI and RNI induced lesions.

Taken together, our data provide evidence that NER in mycobacteria is functional and that UvrB and UvrD play important roles in DNA repair. This is corroborated by analyses of spontaneous mutation frequencies toward rifampicin resistance. *M. smegmatis uvrD* and *uvrB* mutants displayed elevated spontaneous mutation frequencies compared to the wildtype

strain, indicating that disruption of NER has an impact on genomic integrity. Our findings suggest that the substrate spectrum of mycobacterial NER may differ from that of the described model species *E. coli*.

Combined deletions of *uvrB* and *uvrD* have not been analysed in other prokaryotes so far. Interestingly, loss of both proteins rendered *M. smegmatis* more sensitive to UV than deletion of *uvrB* or of *uvrD* alone. The additive effect implies an additional role for one of these proteins besides NER. This assumption is corroborated by the observation that the *in vitro* growth rate of the double mutant was significantly reduced, while growth of the *uvrB* and *uvrD* single mutants was barely affected. We found that no spontaneous resistant clones of the double mutant could be obtained on medium containing high concentrations of rifampicin and that the double mutant was significantly more susceptible to rifampicin as determined by MIC assays. Consequently, the spontaneous mutation frequency towards rifampicin resistance was analysed at lower drug concentrations. The spontaneous mutation frequency of the double mutant was found to be considerably elevated compared to the *uvrB* and *uvrD* single mutants, indicating that overall repair capability is significantly impaired. However, a statistically significant additive effect of combined lack of *uvrB* and *uvrD* was not observed when the strains were assayed for their sensitivity to mitomycin C, TBH and acidified sodium nitrite. Other repair pathways, like BER and recombinational repair, might have overlapping repair specificities for DNA damage generated by these compounds and thus may partially compensate for the lack of *uvrB* and *uvrD* under these conditions.

To investigate mismatch recognition in the absence of a MMR system, we made use of gene conversion assays. The assay applied in this study allows *M. smegmatis* cells to acquire antibiotic resistance upon recombination between the chromosomal *rrnA* operon and a plasmid borne partial gene fragment carrying a specific resistance mutation. This assay provided the opportunity to analyse recognition of different mismatches. Quantification of the relative marker integration frequencies revealed that marker integration depends on the specific mismatch, corresponding to the observations formerly made in MMR proficient species. The efficiency of repair for different base-pairing errors varies and appears to depend on multiple factors such as the physical structure of the mismatch and the sequence context (Dohet *et al.*, 1985; Jones *et al.*, 1987; Marra and Schär, 1999). In *E. coli*, the correction of G/T, A/C, G/G and A/A mismatches by the MMR system is highly effective, while elimination of C/T, T/T and A/G mismatches occurs with decreased efficiency. For C/C

mispairing, no correction activity can be observed (Dohet *et al.*, 1985; Kramer *et al.*, 1984; Parker and Marinus, 1992). We found that the *M. smegmatis uvrB* mutant displayed elevated gene conversion frequencies for most of the mismatches analysed. As gene conversion is mediated by recombination and occurs as consequence of absent mismatch recognition, our results indicate that mycobacterial NER may be involved in limiting recombination and act in the repair of DNA mismatches. A complementary function of NER in recognition of mismatches is corroborated by studies in *Schizosaccharomyces pombe*. NER of *S. pombe* is capable of recognizing and repairing C/C mismatches, which are not substrates of MMR. In the absence of MMR, other mismatches are processed by NER as efficiently as C/C mismatches (Fleck *et al.*, 1999; Kunz and Fleck, 2001). However, we found that marker integration is significantly more efficient in the *M. smegmatis uvrD* mutant than in the *uvrB* mutant. These findings imply that UvrD has a further role besides involvement in the NER pathway. The involvement of mycobacterial UvrD in mismatch recognition and limiting recombination is supported by previous findings made in other prokaryotes. Analysis of *E. coli* UvrD function (Washburn and Kushner, 1991) and more recent studies in *H. pylori* (Kang and Blaser, 2006) revealed that deficiency in UvrD results in a hyperrecombinogenic phenotype. These observations can be explained by two different models. The first model refers to the assumption that NER or MMR remains incomplete as a consequence of UvrD deficiency (Arthur and Lloyd, 1980). Thus, nicks in the DNA occur more frequently. Replication forks will be stalled at such nicks, subsequently leading to the formation of double-strand breaks which will be repaired by recombinational repair (Arthur and Lloyd, 1980). In another model it was proposed that UvrD participates in degradation of toxic recombination intermediates by actively removing RecA proteins from DNA (Centore and Sandler, 2007). This model is based on observations made in yeast and *E. coli*. UvrD of *E. coli* (Centore and Sandler, 2007; Flores *et al.*, 2004; Flores *et al.*, 2005; Petit and Ehrlich, 2002; Veaute *et al.*, 2005) as well as the helicase Srs2 of yeast (Veaute *et al.*, 2003) were found to act as anti-recombinases to prevent potentially deleterious recombination events.

The *uvrB* / *uvrD* double mutant displayed significantly higher gene conversion frequencies than the single mutants. Hence, combined loss of UvrB and UvrD function results in a synergistic (i.e. greater than additive) effect in mycobacteria. The synergism of UvrB and UvrD function in gene conversion prevention was not observed for recognition of A/C or C/C mismatches. The single mutants and the double mutant strain displayed similar marker integration frequencies, which were only marginally elevated compared to the wildtype strain. In the wildtype strain, A/C and C/C gene conversion frequencies are lower than those

obtained with the other mismatches. These findings can be explained by the topological properties of the respective mismatches. A/C and C/C mismatches are the least stable mismatches and lead to significant distortions of the DNA double helix. Hence, due to the unstable nature of the mismatch, the mismatch is rejected irrespective of repair, which accounts for the low marker integration frequencies obtained in the assay. Consequently, only a minor proportion of mismatch intermediate rejection is caused by the anti-recombination function of UvrD.

Taken together, our observations provide evidence that despite the lack of MMR, mycobacteria possess efficient DNA repair mechanisms. We found that mycobacterial NER is capable of repairing a wide range of DNA damage. Furthermore, our data suggest a complementary function of NER in the recognition of DNA mismatches. The helicase component UvrD was shown to limit marker integration frequencies. We propose that repair of DNA mismatches is accomplished differently in mycobacteria than in the described model species. Further investigations will provide insights into new mechanisms that govern genome stability in mycobacteria.

5.3 *M. tuberculosis* NER: implications for pathogenesis and persistence

To analyse whether the observations in *M. smegmatis* could be transferred to the relevant pathogen, *M. tuberculosis* NER mutants were generated. In comparison to the wildtype, the *M. tuberculosis uvrA* mutant and the *uvrA / uvrD* double mutant were found to be more susceptible towards DNA damaging agents, indicating that NER in *M. tuberculosis* is functional. Consistent with the observations made in *M. smegmatis*, the *uvrA* / *uvrD* double mutant displayed a severe growth defect which was accompanied by a significant alteration in colony morphology. These findings point towards an additional role of one of these proteins besides involvement in NER.

The *uvrA /uvrD* double mutant displayed an impaired survival in stationary phase *in vitro* and the *uvrD* mutant exhibited a reduced ability to persist in an *in vivo* mouse model of infection. These findings indicate that besides its function in maintaining genomic stability, NER may also have important roles in the pathogenesis of *M. tuberculosis* infection. In this context it is important to note that the ability to persist for long periods of time within the human host is thought to be the main determinant for the success of *M. tuberculosis* as a human pathogen (Gomez and McKinney, 2004). In fact, it is estimated that one-third of the world population is latently infected and it is assumed that the majority of the disease arises from reactivation of persisting bacteria (Flynn and Chan, 2001). During the latent state, *M. tuberculosis* shifts into a dormant state by shutdown of metabolic activities, including biosynthesis of all cell compounds and replication of the genome. This dormant state is reversible and *M. tuberculosis* cells are able to reactivate when the host immune response is compromised (Flynn and Chan, 2001). The underlying mechanisms to exit dormancy and recover the actively growing state still remain to be elucidated. We hypothesize that during latent infection, mutagenic DNA damage accumulates as the bacteria are under constant attack of DNA damaging agents. Efficient DNA repair systems should thus be a precondition for successful resuscitation of dormant cells. Our analysis of mycobacterial NER corroborates this hypothesis; the *in vivo* phenotype of the *uvrA* / *uvrD* double mutant remains to be determined. Taken together, these findings indicate that functional UvrD is essential for long-term survival of *M. tuberculosis in vivo* and *in vitro* and might thus be critical for emergence from the dormant state during reactivation.

The obtained results may have further implications: the *M. tuberculosis* NER mutants may provide a new avenue for the development of TB vaccines. To date, no vaccine providing efficient protection against tuberculosis in adults is available (Andersen, 2007). The current

tuberculosis vaccine is based on *Mycobacterium bovis* BCG, an attenuated derivative of the causative agent of bovine tuberculosis. However, this vaccine is only capable of protecting children against invasive TB and does not prevent establishment of latent TB or reactivation of pulmonary disease in adults (Arbelaez *et al.*, 2000). With regard to the alerting global impact of tuberculosis, development of novel effective vaccination strategies is of urgent importance. A promising approach for development of a new class of vaccines is based on NER mutants inactivated by photochemical treatment. This concept of so-called killed but metabolically active vaccines (KBMA) was studied using the intracellular pathogen *Listeria monocytogenes* as a model organism (Brockstedt *et al.*, 2005). Deletion of *uvrAB* in *L. monocytogenes* rendered bacteria with increased sensitivity to photochemical inactivation. Thereby growth and the ability to cause disease are prevented. Gene expression, protein synthesis and protein secretion are still proceeding after photochemical inactivation, thus retaining the bacteria's capacity to stimulate innate and adaptive immune responses in the host. As this vaccine is based on a modified form of the bacteria themselves, the authors suppose that this concept might be applied to diseases whose antigenic correlates of immune protection are poorly defined, such as tuberculosis (Brockstedt *et al.*, 2005). This assumption is corroborated by our observations on mycobacterial NER: the *uvrA* and *uvrD* mutants display increased susceptibilities towards irradiation with UV, and functional UvrD is essential for long-term survival of *M. tuberculosis in vitro* and *in vivo*. As growth of the double mutant is severely impaired *in vitro* and presumably attenuated *in vivo*, the double mutant already displays some features of a KBMA, and photochemical inactivation might be not required. Hence, *M. tuberculosis* NER single mutants or the *uvrA* / *uvrD* double mutant in particular might provide a potential basis to develop a KBMA vaccine. Taken together our data provide evidence that the NER system and particularly UvrD play an important role in maintenance of genome stability in mycobacteria, as well as in persistence of *M. tuberculosis*.

5.4 UvrD2 is essential in mycobacteria

The importance of *uvrD* in mycobacteria is corroborated by the existence of a homologue named *uvrD2* which might partially compensate for the loss of *uvrD*. Other Gram-positive bacteria like *Bacillus subtilis* and *Staphylococcus aureus* harbour only one *uvrD* gene (named *pcrA*) which was shown to be essential (Iordanescu, 1993; Petit *et al.*, 1998). *In silico* analyses suggest that mycobacteria and the closely related corynebacteria are the only prokaryotes harbouring a second *uvrD* homologue. In addition to *uvrD2*, mycobacteria possess a third gene (*Rv3201c*) with predicted helicase activity (data obtained by Blastp searches at the TubercuList website of Institute Pasteur). *E. coli* possesses a helicase named Rep in addition to UvrD. Individually, UvrD and Rep proteins are not essential, while the double mutant is not viable (Taucher-Scholz and Hoffmann-Berling, 1983). However, despite homology between UvrD and Rep, the proteins do not share a common activity. It was shown that UvrD directly prevents homologous recombination, whereas Rep acts in replication (Veaute *et al.*, 2005). Moreover, recent studies indicate that *uvrD* defects in *E. coli* cannot be suppressed by Rep, corroborating that UvrD function in *E. coli* is indispensable (Lestini and Michel, 2007).

In an attempt to analyse the function of the two *uvrD* homologues in mycobacteria, we set out to generate *M. tuberculosis* mutants deficient in *uvrD2*, with the aim to subsequently generate double mutants lacking both helicases. However, we were not successful in generating even an *uvrD2* single mutant and therefore assume this homologue to be essential. The reason for lethality remains to be elucidated, as well as the particular function of this helicase.

The identification of conserved essential genes required for growth offers an approach for the development of new antimycobacterial drugs. Thus, *uvrD2* may be considered a promising target for drugs endowed with activity against tuberculosis.

5.5 Role of DNA binding proteins in mycobacteria

Besides DNA repair, other mechanisms that ensure genome integrity are probably important for persistence and resuscitation from the dormant state. DNA binding proteins that establish chromosome architecture presumably contribute to the bacterium's capability to stabilise its genome during persistence and to accomplish proper cell division during reactivation.

We decided to study the role of SMC in mycobacteria as in other organisms this protein was shown to have a global role in maintenance of chromosome stability (Britton *et al.*, 1998; Jensen and Shapiro, 1999; Volkov *et al.*, 2003), as well as in the repair of double-strand breaks (Dervyn *et al.*, 2004). It is assumed that during latency, *M. tuberculosis* experiences an accumulation of double-strand breaks. Functional SMC therefore may be a prerequisite for resuscitation of dormant bacteria. *In silico* analyses revealed that mycobacterial SMC proteins share all structural motifs conserved in members of the SMC family; homologous sequences are restricted to the head, hinge and tail domains and the three domains are connected by two long heptad repeat regions predicted to form coiled-coils.

We successfully generated SMC mutant strains in *M. smegmatis* and *M. tuberculosis*. However, SMC deficiency did not result in an apparent phenotype. This is surprising given previous data obtained in *Bacillus subtilis* and other bacterial SMC mutants, where inactivation of *smc* had a detrimental effect on growth rate and resulted in temperature sensitivity (Britton *et al.*, 1998; Jensen and Shapiro, 1999; Volkov *et al.*, 2003). In contrast, growth of the *M. smegmatis smc* mutant strain was not impaired at any temperature examined. Susceptibility of the mutant to ofloxacin, an inhibitor of DNA gyrase, was comparable to the wildtype. Exposing the mutant to DNA damage caused by ionizing radiation, UV light or treatment with the DNA damaging agent mitomycin C, provided evidence that mycobacterial SMC has only a minor role, if at all, in DNA repair. Our data also indicate that SMC is not essential for resuscitation of dormant bacteria as disruption of SMC did not affect the ability of *M. tuberculosis* and *M. smegmatis* cells to persist in stationary phase.

In addition to SMC, other DNA-binding proteins might participate in maintaining chromosome organisation in mycobacteria and thus compensate for the loss of SMC. Genome sequence analyses indicate that mycobacteria have homologues of many of the proteins found to be associated with the nucleoid in *E. coli*. These proteins include HU (also annotated as HupB, Mdp1 or Hlp1, respectively), integration host factor (IHF) and the histone-like nucleoid structuring protein (H-NS).

Some approaches have been made to characterise the role of histone-like protein HU (*Rv 2986c*) in *M. tuberculosis*. This protein binds to DNA and is capable of suppression of

DNA synthesis, transcription and translation *in vitro* (Furugen *et al.*, 2001). As the expression of HU is increased in stationary growth phase, it might be involved in maintaining genomic stability during persistence of *M. tuberculosis*. However, deletion of the HU homologue in *M. smegmatis* did not result in an apparent phenotype and was not essential for survival in stationary phase, which might be explained by the redundancy of nucleoid-associated proteins in mycobacteria (Lee *et al.*, 1998). This hypothesis is corroborated by the observation that several nucleoid-associated proteins can functionally substitute for each other in *E. coli* (Yasuzawa *et al.*, 1992). A DNA-binding protein named Dps (DNA binding Protein from Starved cells) is present in *M. smegmatis*, but no homologues have been annotated in *M. tuberculosis*. Dps function has been analysed in *E. coli* (Almiron *et al.*, 1992), *Deinococcus radiodurans* (Frenkiel-Krispin *et al.*, 2004), *B. subtilis* (Frenkiel-Krispin and Minsky, 2006), and *M. smegmatis* (Gupta *et al.*, 2002; Gupta and Chatterji, 2003). The protein is upregulated under starvation conditions and binds DNA without apparent sequence specificity, forming extremely stable complexes. In stationary phase, abundant Dps proteins protect DNA through Dps–DNA co-crystallization (Wolf *et al.*, 1999). This protein might be the main substitute for HU function in *M. smegmatis.* It remains to be shown whether the findings on HU obtained in the model organism *M. smegmatis* hold true for its pathogenic relative *M. tuberculosis*.

Chromosome structure is also linked to the regulation of gene expression and many nucleoid-associated proteins have been found to act as transcription factors besides their function as structural proteins (Stavans and Oppenheim, 2006). Recent studies of the *iniBAC* genes that confer antibiotic tolerance in *M. tuberculosis* revealed that those genes are induced by a histone-like protein named *lsr2* (Colangeli *et al.*, 2007). These data indicate that nucleoid – associated proteins might have an important role in the pathogenesis of *M. tuberculosis*.

5.6 Conclusions

With this work we made important progress towards understanding the mechanisms involved in maintenance of genome integrity in mycobacteria. Analysis of mycobacterial NER provided new insights into mycobacterial DNA repair by pointing out that mycobacteria significantly differ in this regard from described model species. Our findings imply that UvrD has an important role in compensating for MMR deficiency and that UvrD is involved in *M. tuberculosis* persistence.

While the *M. tuberculosis* NER mutants may offer a promising avenue for the development of a new class of TB vaccines, further analysis of the mechanisms that govern chromatin stability in *M. tuberculosis* may lead to the development of new drugs active against bacteria in the latent state. These antibiotics could be of general use both to shorten the time period required to treat active disease and to eradicate latent tuberculosis.

6 References

Almiron, M., Link, A.J., Furlong, D., and Kolter, R. (1992) A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev* **6**: 2646-2654.

Andersen, P. (2007) Tuberculosis vaccines - an update. *Nat Rev Microbiol* **5**: 484-487.

Arbelaez, M.P., Nelson, K.E., and Munoz, A. (2000) BCG vaccine effectiveness in preventing tuberculosis and its interaction with human immunodeficiency virus infection. *Int J Epidemiol* **29**: 1085-1091.

Arthur, H.M., and Lloyd, R.G. (1980) Hyper-recombination in *uvrD* mutants of *Escherichia coli* K-12. *Mol Gen Genet* **180**: 185-191.

Batty, D.P., and Wood, R.D. (2000) Damage recognition in nucleotide excision repair of DNA. *Gene* **241**: 193-204.

Böttger, E.C. (1994) Resistance to drugs targeting protein synthesis in mycobacteria. *Trends Microbiol* **2**: 416-421.

Böttger, E.C. (2006) Mechanisms of mycobacterial drug resistance: implications for treatment and prevention strategies. *Chemotherapie J* **15**: 23-28.

Böttger, E.C., and Springer, B. (2008) Tuberculosis: drug resistance, fitness, and strategies for global control. *Eur J Pediatr* **167**: 141-148.

Britton, R.A., Lin, D.C.-H., and Grossman, A.D. (1998) Characterization of a prokaryotic SMC protein involved in chromosome partitioning. *Genes Dev* **12**: 1254-1259.

Brockstedt, D.G., Bahjat, K.S., Giedlin, M.A., Liu, W., Leong, M., Luckett, W., Gao, Y., Schnupf, P., Kapadia, D., Castro, G., Lim, J.Y., Sampson-Johannes, A., Herskovits, A.A., Stassinopoulos, A., Bouwer, H.G., Hearst, J.E., Portnoy, D.A., Cook, D.N., and Dubensky, T.W., Jr. (2005) Killed but metabolically active microbes: a new vaccine paradigm for eliciting effector T-cell responses and protective immunity. *Nat Med* **11**: 853-860.

Brosch, R., Gordon, S.V., Marmiesse, M., Brodin, P., Buchrieser, C., Eiglmeier, K., Garnier, T., Gutierrez, C., Hewinson, G., Kremer, K., Parsons, L.M., Pym, A.S., Samper, S., van Soolingen, D., and Cole, S.T. (2002) A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc Natl Acad Sci U S A* **99**: 3684-3689.

Burma, S., Chen, B.P., and Chen, D.J. (2006) Role of non-homologous end joining (NHEJ) in maintaining genomic integrity. *DNA Repair (Amst)* **5**: 1042-1048.

Burney, S., Caulfield, J.L., Niles, J.C., Wishnok, J.S., and Tannenbaum, S.R. (1999) The chemistry of DNA damage from nitric oxide and peroxynitrite. *Mutat Res* **424**: 37-49.

Centore, R.C., and Sandler, S.J. (2007) UvrD limits the number and intensities of RecA-green fluorescent protein structures in *Escherichia coli* K-12. *J Bacteriol* **189**: 2915-2920.

Champoux, J.J. (2001) DNA topoisomerases: structure, function and mechanism. *Annu Rev Biochem* **70**: 369-413.

Chan, J., Xing, Y., Magliozzo, R.S., and Bloom, B.R. (1992) Killing of virulent *Mycobacterium tuberculosis* by reactive nitrogen intermediates produced by activated murine macrophages. *J Exp Med* **175**: 1111-1122.

Cobbe, N., and Heck, M.M.S. (2000) Review: SMCs in the world of chromosome biology-- from prokaryotes to higher eukaryotes. *J Struct Biol* **129**: 123-143.

Cobbe, N., and Heck, M.M. (2004) The evolution of SMC proteins: phylogenetic analysis and structural implications. *Mol Biol Evol* **21**: 332-347.

Colangeli, R., Helb, D., Vilcheze, C., Hazbon, M.H., Lee, C.-G., Jr, W.R.J., and Alland, D. (2007) Multi-drug resistant tolerance is transcriptionally regulated by the histone-like protein Lsr2 in *Mycobacterium tuberculosis*. In *Keystone Symposia: Tuberculosis: From Lab Research to Field Trials, Abstr 157* Vancouver, Canada.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., and Barrell, B.G. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007-1011.

Cox, M.M., Goodman, M.F., Kreuzer, K.N., Sherratt, D.J., Sandler, S.J., and Marians, K.J. (2000) The importance of repairing stalled replication forks. *Nature* **404**: 37-41.

Darwin, K.H., and Nathan, C.F. (2005) Role for nucleotide excision repair in virulence of *Mycobacterium tuberculosis*. *Infect Immun* **73**: 4581-4587.

David, S.S., O'Shea, V.L., and Kundu, S. (2007) Base-excision repair of oxidative DNA damage. *Nature* **447**: 941-950.

Davis, E.O., Springer, B., Gopaul, K.K., Papavinasasundaram, K.G., Sander, P., and Böttger, E.C. (2002) DNA damage induction of *recA* in *Mycobacterium tuberculosis* independently of RecA and LexA. *Mol Microbiol* **46**: 791-800.

Demple, B., and Harrison, L. (1994) Repair of oxidative damage to DNA: enzymology and biology. *Annu Rev Biochem* **63**: 915-948.

Dervyn, E., Noirot-Gros, M.F., Mervelet, P., McGovern, S., Ehrlich, S.D., Polard, P., and Noirot, P. (2004) The bacterial condensin/cohesin-like protein complex acts in DNA repair and regulation of gene expression. *Mol Microbiol* **51**: 1629-1640.

Ding, A., Nathan, C., and Stuehr, D. (1988) Release of reactive nitrogen intermediates and reactive oxygen intermediates from mouse peritoneal macrophages. Comparison of activating cytokines and evidence for independent production. *J Immunol* **141**: 2407-2412.

Dohet, C., Wagner, R., and Radman, M. (1985) Repair of defined single base-pair mismatches in *Escherichia coli*. *Proc Natl Acad Sci U S A* **82**: 503-505.

Dye, C., Watt, C.J., Bleed, D.M., Hosseini, S.M., and Raviglione, M.C. (2005) Evolution of tuberculosis control and prospects for reducing tuberculosis incidence, prevalence, and deaths globally. *Jama* **293**: 2767-2775.

Finken, M., Kirschner, P., Meier, A., Wrede, A., and Böttger, E.C. (1993) Molecular basis of streptomycin resistance in *Mycobacterium tuberculosis*: alterations of the ribosomal protein S12 gene and point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol Microbiol* **9**: 1239-1246.

Fleck, O., Lehmann, E., Schar, P., and Kohli, J. (1999) Involvement of nucleotide-excision repair in *msh2 pms1*-independent mismatch repair. *Nat Genet* **21**: 314-317.

Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O., Peterson, J., DeBoy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E., Kolonay, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M., Salzberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H., Gill, J., Mikula, A., Bishai, W., Jacobs Jr, W.R., Jr., Venter, J.C., and Fraser, C.M. (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J Bacteriol* **184**: 5479-5490.

Flores, M.J., Bidnenko, V., and Michel, B. (2004) The DNA repair helicase UvrD is essential for replication fork reversal in replication mutants. *EMBO Rep* **5**: 983-988.

Flores, M.J., Sanchez, N., and Michel, B. (2005) A fork-clearing role for UvrD. *Mol Microbiol* **57**: 1664-1675.

Flynn, J.L., Scanga, C.A., Tanaka, K.E., and Chan, J. (1998) Effects of aminoguanidine on latent murine tuberculosis. *J Immunol* **160**: 1796-1803.

Flynn, J.L., and Chan, J. (2001) Tuberculosis: latency and reactivation. *Infect Immun* **69**: 4195-4201.

Frenkiel-Krispin, D., Ben-Avraham, I., Englander, J., Shimoni, E., Wolf, S.G., and Minsky, A. (2004) Nucleoid restructuring in stationary-state bacteria. *Mol Microbiol* **51**: 395-405.

Frenkiel-Krispin, D., and Minsky, A. (2006) Nucleoid organization and the maintenance of DNA integrity in *E. coli*, *B. subtilis* and *D. radiodurans*. *J Struct Biol* **156**: 311-319.

Friedberg, E.C., Graham C. Walker, Wolfram Siede, Richard D. Wood, Roger A. Schultz, and Tom Ellenberger (2005) *DNA Repair and Mutagenesis.* Washington DC: American Society of Microbiology Press.

Frischkorn, K., Sander, P., Scholz, M., Teschner, K., Prammananan, T., and Böttger, E.C. (1998) Investigation of mycobacterial *recA* function: protein introns in the RecA of pathogenic mycobacteria do not affect competency for homologous recombination. *Mol Microbiol* **29**: 1203-1214.

Furin, J. (2007) The clinical management of drug-resistant tuberculosis. *Curr Opin Pulm Med* **13**: 212-217.

Furugen, M., Matsumoto, S., Matsuo, T., Matsumoto, M., and Yamada, T. (2001) Identification of the mycobacterial DNA-binding protein 1 region which suppresses transcription *in vitro*. *Microb Pathog* **30**: 129-138.

Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V., and Hewinson, R.G. (2003) The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A* **100**: 7877-7882.

Gomez, J.E., and McKinney, J.D. (2004) *M. tuberculosis* persistence, latency, and drug tolerance. *Tuberculosis* **84**: 29-44.

Gong, C., Bongiorno, P., Martins, A., Stephanou, N.C., Zhu, H., Shuman, S., and Glickman, M.S. (2005) Mechanism of nonhomologous end-joining in mycobacteria: a low-fidelity repair system driven by Ku, ligase D and ligase C. *Nat Struct Mol Biol* **12**: 304-312.

Goodman, M.F. (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* **71**: 17-50.

Gupta, S., Pandit, S.B., Srinivasan, N., and Chatterji, D. (2002) Proteomics analysis of carbon-starved *Mycobacterium smegmatis*: induction of Dps-like protein. *Protein Eng* **15**: 503-512.

Gupta, S., and Chatterji, D. (2003) Bimodal protection of DNA by M*ycobacterium smegmatis* DNA-binding protein from stationary phase cells. *J Biol Chem* **278**: 5235-5241.

Halliwell, B., and Aruoma, O.I. (1991) DNA damage by oxygen-derived species. Its mechanism and measurement in mammalian systems. *FEBS Lett* **281**: 9-19.

Hirano, M., and Hirano, T. (2004) Positive and negative regulation of SMC-DNA interactions by ATP and accessory proteins. *EMBO J* **23**: 2664-2673.

Hirano, T. (2005) SMC proteins and chromosome mechanics: from bacteria to humans. *Philos Trans R Soc Lond B Biol Sci* **360**: 507-514.

Hirsh, A.E., Tsolaki, A.G., DeRiemer, K., Feldman, M.W., and Small, P.M. (2004) Stable association between strains of *Mycobacterium tuberculosis* and their human host populations. *Proc Natl Acad Sci U S A* **101**: 4871-4876.

Hix, S., Morais Mda, S., and Augusto, O. (1995) DNA methylation by tert-butyl hydroperoxide-iron (II). *Free Radic Biol Med* **19**: 293-301.

Huffman, J.L., Sundheim, O., and Tainer, J.A. (2005) DNA base damage recognition and removal: New twists and grooves. *Mutat Res* **577**: 55-76.

Imlay, J.A., and Linn, S. (1987) Mutagenesis and stress responses induced in *Escherichia coli* by hydrogen peroxide. *J Bacteriol* **169**: 2967-2976.

Iordanescu, S. (1993) Characterization of the *Staphylococcus aureus* chromosomal gene *pcrA*, identified by mutations affecting plasmid pT181 replication. *Mol Gen Genet* **241**: 185-192.

Ishino, Y., Nishino, T., and Morikawa, K. (2006) Mechanisms of maintaining genetic stability by homologous recombination. *Chem Rev* **106**: 324-339.

Iyer, R.R., Pluciennik, A., Burdett, V., and Modrich, P.L. (2006) DNA mismatch repair: Functions and mechanisms. *Chem Rev* **106**: 302-323.

Iyer, V.N., and Szybalski, W. (1963) A Molecular Mechanism of Mitomycin Action: Linking of Complementary DNA Strands. *Proc Natl Acad Sci U S A* **50**: 355-362.

Jensen, R.B., and Shapiro, L. (1999) The *Caulobacter crescentus smc* gene is required for cell cycle progression and chromosome segregation. *Proc Natl Acad Sci U S A* **96**: 10661-10666.

Jessberger, R. (2002) The many functions of SMC proteins in chromosome dynamics. *Nature Rev Mol Cell Biol* **3**: 767-778.

Jones, M., Wagner, R., and Radman, M. (1987) Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* **115**: 605-610.

Kang, J., and Blaser, M.J. (2006) UvrD helicase suppresses recombination and DNA damage-induced deletions. *J Bacteriol* **188**: 5450-5459.

Kaufmann, S.H. (2001) How can immunology contribute to the control of tuberculosis? *Nat Rev Immunol* **1**: 20-30.

Kenney, T.J., and Churchward, G. (1994) Cloning and sequence analysis of the *rpsL* and *rpsG* genes of *Mycobacterium smegmatis* and characterization of mutations causing resistance to streptomycin. *J Bacteriol* **176**: 6153-6156.

Konola, J.T., Sargent, K.E., and Gow, J.B. (2000) Efficient repair of hydrogen peroxide-induced DNA damage by *Escherichia coli* requires SOS induction of RecA and RuvA proteins. *Mutat Res* **459**: 187-194.

Kramer, B., Kramer, W., and Fritz, H.J. (1984) Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. *Cell* **38**: 879-887.

Kumar, S., Lipman, R., and Tomasz, M. (1992) Recognition of specific DNA sequences by mitomycin C for alkylation. *Biochemistry* **31**: 1399-1407.

Kunkel, T.A., and Erie, D.A. (2005) DNA mismatch repair. *Annu Rev Biochem* **74**: 681-710.

Kunz, C., and Fleck, O. (2001) Role of the DNA repair nucleases Rad13, Rad2 and Uve1 of *Schizosaccharomyces pombe* in mismatch correction. *J Mol Biol* **313**: 241-253.

Labes, M., Puhler, A., and Simon, R. (1990) A new family of RSF1010-derived expression and lac-fusion broad-host-range vectors for Gram-negative bacteria. *Gene* **89**: 37-46.

Lea, D.E., and Coulson, C.A. (1949) The distribution of the numbers of mutants in bacterial populations. *J Gen* **49**: 264-285.

LeClerc, J.E., Li, B., Payne, W.L., and Cebula, T.A. (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science* **274**: 1208-1211.

Lederberg, J. (1951) Streptomycin resistance; a genetically recessive mutation. *J Bacteriol* **61**: 549-550.

Lee, B.H., Murugasu-Oei, B., and Dick, T. (1998) Upregulation of a histone-like protein in dormant *Mycobacterium smegmatis*. *Mol Gen Genet* **260**: 475-479.

Lehmann, A.R., Walicka, M., Griffiths, D.J., Murray, J.M., Watts, F.Z., McCready, S., and Carr, A.M. (1995) The rad18 gene of *Schizosaccharomyces pombe* defines a new subgroup of the SMC superfamily involved in DNA repair. *Mol Cell Biol* **15**: 7067-7080.

Lehmann, A.R. (2005) The role of SMC proteins in the responses to DNA damage. *DNA Repair* **4**: 309-314.

Lestini, R., and Michel, B. (2007) UvrD controls the access of recombination proteins to blocked replication forks. *Embo J*.

Losada, A., and Hirano, T. (2005) Dynamic molecular linkers of the genome: the first decade of SMC proteins. *Genes Dev.* **19**: 1269-1287.

Luria, S.E., and Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491-511.

Mariam, D.H., Mengistu, Y., Hoffner, S.E., and Andersson, D.I. (2004) Effect of *rpoB* mutations conferring rifampin resistance on fitness of *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **48**: 1289-1294.

Marra, G., and Schär, P. (1999) Recognition of DNA alterations by the mismatch repair system. *Biochem J* **338 (Pt 1)**: 1-13.

Matic, I., Rayssiguier, C., and Radman, M. (1995) Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. *Cell* **80**: 507-515.

Meier, A., Sander, P., Schaper, K.J., Scholz, M., and Böttger, E.C. (1996) Correlation of molecular resistance mechanisms and phenotypic resistance levels in streptomycin-resistant *Mycobacterium tuberculosis*. *Antimicrob Agents Chemother* **40**: 2452-2454.

Minton, K.W. (1994) DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. *Mol Microbiol* **13**: 9-15.

Mishina, Y., Duguid, E.M., and He, C. (2006) Direct reversal of DNA alkylation damage. *Chem Rev* **106**: 215-232.

Mizrahi, V., and Andersen, S.J. (1998) DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Mol Microbiol* **29**: 1331-1339.

Musser, J. (1995) Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clin Microbiol Rev* **8**: 496-514.

Nathan, C., and Shiloh, M.U. (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. *Proc Natl Acad Sci U S A* **97**: 8841-8848.

Nohmi, T. (2006) Environmental stress and lesion-bypass DNA polymerases. *Annu Rev Microbiol* **60**: 231-253.

Nowosielska, A., Smith, S.A., Engelward, B.P., and Marinus, M.G. (2006) Homologous recombination prevents methylation-induced toxicity in *Escherichia coli*. *Nucleic Acids Res* **34**: 2258-2268.

Parker, B.O., and Marinus, M.G. (1992) Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc Natl Acad Sci U S A* **89**: 1730-1734.

Petit, M.A., Dervyn, E., Rose, M., Entian, K.D., McGovern, S., Ehrlich, S.D., and Bruand, C. (1998) PcrA is an essential DNA helicase of *Bacillus subtilis* fulfilling functions both in repair and rolling-circle replication. *Mol Microbiol* **29**: 261-273.

Petit, M.A., and Ehrlich, D. (2002) Essential bacterial helicases that counteract the toxicity of recombination proteins. *EMBO J* **21**: 3137-3147.

Pfister, P., Hobbie, S., Vicens, Q., Böttger, E.C., and Westhof, E. (2003) The molecular basis for A-Site mutations conferring aminoglycoside resistance: Relationship between ribosomal susceptibility and x-ray crystal structures. *ChemBioChem* **4**: 1078-1088.

Pfister, P., Jenni, S., Poehlsgaard, J., Thomas, A., Douthwaite, S., Ban, N., and Böttger, E.C. (2004) The structural basis of macrolide-ribosome binding assessed using mutagenesis of 23 S rRNA positions 2058 and 2059. *J Mol Biol* **342**: 1569-1581.

Pfister, P., Hobbie, S., Brüll, C., Corti, N., Vasella, A., Westhof, E., and Böttger, E.C. (2005) Mutagenesis of 16 S *rRNA* C1409-G1491 base-pair differentiates between 6'OH and 6'NH3+ aminoglycosides. *J Mol Biol* **346**: 467-475.

Pitcher, R.S., Green, A.J., Brzostek, A., Korycka-Machala, M., Dziadek, J., and Doherty, A.J. (2007) NHEJ protects mycobacteria in stationary phase against the harmful effects of desiccation. *DNA Repair*.

Potts, P.R., Porteus, M.H., and Yu, H. (2006) Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. *EMBO J* **25**: 3377-3388.

Ramaswamy, S., and Musser, J.M. (1998) Molecular genetic basis of antimicrobial agent resistance in *Mycobacterium tuberculosis*: 1998 update. *Tuber Lung Dis* **79**: 3-29.

Rand, L., Hinds, J., Springer, B., Sander, P., Buxton, R.S., and Davis, E.O. (2003) The majority of inducible DNA repair genes in *Mycobacterium tuberculosis* are induced independently of *RecA*. *Mol Microbiol* **50**: 1031-1042.

Raynaud, C., Papavinasasundaram, K.G., Speight, R.A., Springer, B., Sander, P., Böttger, E.C., Colston, M.J., and Draper, P. (2002) The functions of OmpATb, a pore-forming protein of *Mycobacterium tuberculosis*. *Mol Microbiol* **46**: 191-201.

Rayssiguier, C., Thaler, D.S., and Radman, M. (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* **342**: 396-401.

Rezwan, M. (2006) Lipoprotein maturation is required for virulence of *Mycobacterium tuberculosis* and is essential for correct lipoprotein localization. Zurich: Swiss Federal Institute of Technology Zurich (PhD Thesis).

Sambrook, J., E. F. Fritsch, and T. Maniatis (1989) *Molecular cloning: a laboratory manual.* Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Samson, L., Thomale, J., and Rajewsky, M.F. (1988) Alternative pathways for the *in vivo* repair of O6-alkylguanine and O4-alkylthymine in *Escherichia coli*: the adaptive response and nucleotide excision repair. *Embo J* **7**: 2261-2267.

Sancar, A. (1996) DNA Excision Repair. *Annu Rev Biochem* **65**: 43-81.

Sancar, A. (2003) Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors. *Chem Rev* **103**: 2203-2237.

Sander, P., Meier, A., and Böttger, E.C. (1995) *rpsL*+: a dominant selectable marker for gene replacement in mycobacteria. *Mol Microbiol* **16**: 991-1000.

Sander, P., Prammananan, T., and Böttger, E.C. (1996) Introducing mutations into a chromosomal *rRNA* gene using a genetically modified eubacterial host with a single *rRNA* operon. *Mol Microbiol* **22**: 841-848.

Sander, P., Papavinasasundaram, K.G., Dick, T., Stavropoulos, E., Ellrott, K., Springer, B., Colston, M.J., and Böttger, E.C. (2001a) *Mycobacterium bovis* BCG *recA* deletion mutant shows increased susceptibility to DNA-damaging agents but wild-type survival in a mouse infection model. *Infect Immun* **69**: 3562–3568.

Sander, P., Springer, B., and Böttger, E.C. (2001b) Gene replacement in *Mycobacterium tuberculosis* and *Mycobacterium bovis* BCG using *rpsL+* as a dominant negative selectable marker. In *Methods in Molecular Medicine: Mycobacterium tuberculosis Protocols.* Parish T. and Stoker, N.G. (eds). Totowa, New Jersey: Humana Press, pp. 93-104.

Sassetti, C.M., Boyd, D.H., and Rubin, E.J. (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol Microbiol* **48**: 77-84.

Sassetti, C.M., and Rubin, E.J. (2003) Genetic requirements for mycobacterial survival during infection. *Proc Natl Acad Sci U S A* **100**: 12989-12994.

Sawitzke, J.A., and Austin, S. (2000) Suppression of chromosome segregation defects of *Escherichia coli muk* mutants by mutations in topoisomerase I. *Proc Natl Acad Sci U S A* **97**: 1671-1676.

Schlacher, K., Pham, P., Cox, M.M., and Goodman, M.F. (2006) Roles of DNA polymerase V and RecA protein in SOS damage-induced mutation. *Chem Rev* **106**: 406-419.

Schlagman, S.L., Hattman, S., and Marinus, M.G. (1986) Direct role of the *Escherichia coli* Dam DNA methyltransferase in methylation-directed mismatch repair. *J Bacteriol* **165**: 896-900.

Schofield, M.J., and Hsieh, P. (2003) DNA mismatch repair: molecular mechanisms and biological function. *Annu Rev Microbiol* **57**: 579-608.

Shah, N., Wright, A., Bai, G.-H., Barrera, L., Boulahbal, F., and Martín-Casabona, N. (2007) Worldwide emergence of extensively drug-resistant tuberculosis. In *Emerg Infect Dis [serial on the Internet]. Available from* http://www.cdc.gov/EID/content/13/3/380.htm.

Sicard, N., Oreglia, J., and Estevenon, A.M. (1992) Structure of the gene complementing uvr-402 in *Streptococcus pneumoniae*: homology with *Escherichia coli uvrB* and the homologous gene in *Micrococcus luteus*. *J Bacteriol* **174**: 2412-2415.

Sidorkina, O., Saparbaev, M., and Laval, J. (1997) Effects of nitrous acid treatment on the survival and mutagenesis of *Escherichia coli* cells lacking base excision repair (hypoxanthine-DNA glycosylase-ALK A protein) and/or nucleotide excision repair. *Mutagenesis* **12**: 23-28.

Sjogren, C., and Nasmyth, K. (2001) Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. *Curr Biol* **11**: 991-995.

Soppa, J. (2001) Prokaryotic structural maintenance of chromosomes (SMC) proteins: distribution, phylogeny, and comparison with MukBs and additional prokaryotic and eukaryotic coiled-coil proteins. *Gene* **278**: 253-264.

Spek, E.J., Wright, T.L., Stitt, M.S., Taghizadeh, N.R., Tannenbaum, S.R., Marinus, M.G., and Engelward, B.P. (2001) Recombinational repair is critical for survival of *Escherichia coli* exposed to nitric oxide. *J Bacteriol* **183**: 131-138.

Springer, B., Sander, P., Sedlacek, L., Hardt, W.-D., Mizrahi, V., Schar, P., and Böttger, E.C. (2004) Lack of mismatch correction facilitates genome evolution in mycobacteria. *Mol Microbiol* **53**: 1601-1609.

Stavans, J., and Oppenheim, A. (2006) DNA-protein interactions and bacterial chromosome architecture. *Phys Biol* **3**: R1-10.

Stephanou, N.C., Gao, F., Bongiorno, P., Ehrt, S., Schnappinger, D., Shuman, S., and Glickman, M.S. (2007) Mycobacterial nonhomologous end joining mediates mutagenic repair of chromosomal double-strand DNA breaks. *J Bacteriol* **189**: 5237-5246.

Stover, C.K., de la Cruz, V.F., Fuerst, T.R., Burlein, J.E., Benson, L.A., Bennett, L.T., Bansal, G.P., Young, J.F., Lee, M.H., Hatfull, G.F., and et al. (1991) New use of BCG for recombinant vaccines. *Nature* **351**: 456-460.

Tadesse, S., and Graumann, P.L. (2006) Differential and dynamic localization of topoisomerases in *Bacillus subtilis*. *J Bacteriol* **188**: 3002-3011.

Taucher-Scholz, G., and Hoffmann-Berling, H. (1983) Identification of the gene for DNA helicase II of *Escherichia coli*. *Eur J Biochem* **137**: 573-580.

Van Houten, B., Croteau, D.L., DellaVecchia, M.J., Wang, H., and Kisker, C. (2005) 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutat Res* **577**: 92-117.

Veaute, X., Jeusset, J., Soustelle, C., Kowalczykowski, S.C., Le Cam, E., and Fabre, F. (2003) The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**: 309-312.

Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M.A. (2005) UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in *Escherichia coli*. *EMBO J* **24**: 180-189.

Venkatesh, J., Kumar, P., Krishna, P.S., Manjunath, R., and Varshney, U. (2003) Importance of uracil DNA glycosylase in *Pseudomonas aeruginosa* and *Mycobacterium smegmatis*, G+C-rich bacteria, in mutation prevention, tolerance to acidified nitrite, and endurance in mouse macrophages. *J Biol Chem* **278**: 24350-24358.

Vidal, L.S., Santos, L.B., Lage, C., and Leitao, A.C. (2006) Enhanced sensitivity of *Escherichia coli uvrB* mutants to mitomycin C points to a UV-C distinct repair for DNA adducts. *Chem Res Toxicol* **19**: 1351-1356.

Volkov, A., Mascarenhas, J., Andrei-Selmer, C., Ulrich, H.D., and Graumann, P.L. (2003) A prokaryotic condensin/cohesin-like complex can actively compact chromosomes from a single position on the nucleoid and binds to DNA as a ring-like structure. *Mol Cell Biol* **23**: 5638-5650.

Warner, D.F., and Mizrahi, V. (2006) Tuberculosis chemotherapy: the influence of bacillary stress and damage response pathways on drug efficacy. *Clin Microbiol Rev* **19**: 558-570.

Washburn, B.K., and Kushner, S.R. (1991) Construction and analysis of deletions in the structural gene (*uvrD*) for DNA helicase II of *Escherichia coli*. *J Bacteriol* **173**: 2569-2575.

Wayne, L.G., and Hayes, L.G. (1996) An *in vitro* model for sequential study of shiftdown of *Mycobacterium tuberculosis* through two stages of nonreplicating persistence. *Infect Immun* **64**: 2062-2069.

Wolf, S.G., Frenkiel, D., Arad, T., Finkel, S.E., Kolter, R., and Minsky, A. (1999) DNA protection by stress-induced biocrystallization. *Nature* **400**: 83-85.

Worth, L., Jr., Clark, S., Radman, M., and Modrich, P. (1994) Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc Natl Acad Sci U S A* **91**: 3238-3241.

Wyman, C., Ristic, D., and Kanaar, R. (2004) Homologous recombination-mediated double-strand break repair. *DNA Repair (Amst)* **3**: 827-833.

Yasuzawa, K., Hayashi, N., Goshima, N., Kohno, K., Imamoto, F., and Kano, Y. (1992) Histone-like proteins are required for cell growth and constraint of supercoils in DNA. *Gene* **122**: 9-15.

Zahrt, T.C., and Deretic, V. (2002) Reactive Nitrogen and Oxygen Intermediates and Bacterial Defenses: Unusual Adaptations in *Mycobacterium tuberculosis*. *Antioxidants & Redox Signaling* **4**: 141-159.

7 Appendix

7.1 Curriculum vitae

Personal data:

Name:	GÜTHLEIN
First name:	Carolin
Date of birth:	13 December 1978
Nationality:	German

Education:

09/1989 – 05/1998	Gutenberg-Gymnasium Wiesbaden, Germany **Abitur (Qualification for university entrance)**
10/1998 -09/2003	Study of Biology at Johann-Wolfgang-Goethe University, Frankfurt am Main, Germany **Graduated as Diplom-Biologin**
11/2002	**Diploma examinations**: Genetics, Microbiology and Biochemistry
12/2002- 09/2003	**Diploma thesis** within the working group of Prof. K.-D. Entian at the Institute of Microbiology, University of Frankfurt, Germany. Subject: "Analysis of co-activator interactions in *Saccharomyces cerevisiae*"
Since 12/2003	**PhD thesis** at the Institute of Medical Microbiology, University Zurich under supervision of Prof. E.C. Böttger

Publications:

"A mycobacterial *smc* null mutant is proficient in DNA repair and long-term survival" (2008)
C. Güthlein, R. M. Wanner, P. Sander, E. C. Böttger and B. Springer
Journal of Bacteriology Vol. 190, 452-456

"Characterisation of mycobacterial NER reveals accessory functions of the helicase *uvrD*"
C. Güthlein, R. M. Wanner, P. Sander, M. Bosshard, E. C. Böttger and B. Springer
Journal of Bacteriology (submitted)

"Stabilization of the genome of the mismatch repair deficient *M. tuberculosis* by context-dependent codon choice"
R. M. Wanner, C. Güthlein, E.C. Böttger, B. Springer and M. Ackermann
BMC Genomics (submitted)

7.2 Publications

Characterisation of the mycobacterial NER system reveals accessory functions of the helicase *uvrD*

Carolin Güthlein[1], Roger M. Wanner[1], Peter Sander[1,2], Elaine O. Davis[3], Martin Bosshard[1], Erik C. Böttger[1,2] and Burkhard Springer[1,2,4]

[1] Institut für Medizinische Mikrobiologie, Universität Zürich, Gloriastr. 30/32, CH-8006 Zürich, Schweiz

[2] Nationales Zentrum für Mykobakterien, Gloriastr. 30, CH-8006 Zürich, Schweiz

[3] Division of Mycobacterial Research, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

[4] current address: Institut für Medizinische Mikrobiologie und Hygiene, Agentur für Gesundheit und Ernährungssicherheit, Beethovenstrasse 6, 8010 Graz, Österreich

Abstract

In this study, we investigated the role of the nucleotide excision repair (NER) pathway in mycobacteria, microorganisms naturally devoid of the mismatch repair (MMR) system. Single knock-out mutants deficient in the excinuclease component *uvrB* and in the helicase *uvrD* and a double knock-out lacking both proteins were constructed in *Mycobacterium smegmatis.* Studying susceptibility to DNA damaging agents, we found that the mycobacterial NER system is involved in the repair of a wide range of mutagenic DNA lesions and that the mycobacterial NER system is an important defence mechanism against oxidative and nitrosative damage. In a gene conversion assay, recognition and subsequent rejection of base-pairing errors was assessed. Inactivation of *uvrB* and *uvrD* increased marker integration frequencies, with in part – and dependent on the mismatch studied - synergistic effects in the combined *uvrB* / *uvrD* mutant. Our results imply that NER and particularly the helicase UvrD have important roles in maintaining genome integrity and contribute to compensating for the lack of MMR in mycobacteria.

Introduction

Part of the success of the pathogen *Mycobacterium tuberculosis* lies in its ability to survive and replicate in macrophages, although the molecular mechanisms allowing it to do so are not well understood (Kaufmann, 2001). The production of reactive oxygen intermediates (ROI) and reactive nitrogen (RNI) intermediates by macrophages is considered an effective defence mechanism against intracellular pathogens (Chan *et al.*, 1992; Ding *et al.*, 1988; Nathan and Shiloh, 2000). Among other cellular targets, ROI and RNI attack the chemical compounds of the DNA, i.e. the sugar-phosphate backbone and the bases (Burney *et al.*, 1999). *M. tuberculosis* seems to be superbly adapted to these unfavourable conditions which demand efficient mechanisms that ensure genome stability. Even though efforts have been made to gain insight into mycobacterial DNA repair processes (Curti *et al.*, 2007; Durbach *et al.*, 2003; Sander *et al.*, 2001; Springer *et al.*, 2001), the mechanisms involved in maintenance of genomic fidelity remain ambiguous. *In silico* analyses of mycobacterial genomes, i.e. *M. tuberculosis* (Cole *et al.*, 1998; Mizrahi and Andersen, 1998), *M. leprae* (Cole *et al.*, 2001), *M. bovis* (Garnier *et al.*, 2003), *M. avium, M. paratuberculosis* and *M. smegmatis* (The Institute for Genome Research; http://www.tigr.org) revealed the presence of genes coding for enzymes involved in damage reversal, nucleotide excision repair (NER), base excision repair (BER), recombinational repair, non-homologous end-joining (NHEJ) and SOS repair. However, mycobacteria are devoid of the *mutLS*-based postreplicative mismatch repair (MMR) system (Mizrahi and Andersen, 1998; Springer *et al.*, 2004). The only MMR homologue present in mycobacteria is a DNA-helicase named UvrD, which is also part of the NER. The MMR system is highly conserved throughout evolution and contributes to mutation avoidance by correcting replication errors resulting from nucleotide misincorporation and polymerase slippage (Kunkel and Erie, 2005; Schofield and Hsieh, 2003). In addition, MMR inhibits recombination between non-identical (homeologous) sequences and thus takes part in ensuring the fidelity of recombination (Matic *et al.*, 1995; Rayssiguier *et al.*, 1989; Worth *et al.*, 1994).

The finding that mycobacteria exhibit a general mutation rate that is comparable to that of MMR proficient species, suggests that mycobacteria possess alternative or compensating strategies for mismatch recognition and repair (Springer *et al.*, 2004). Candidates for compensating mechanisms are known DNA repair processes, e.g. NER. Prokaryotic NER has been extensively analysed in *E. coli*. The process of NER in bacteria is mediated by the UvrABC excinuclease enzyme complex and the helicase UvrD (Sancar, 1996). The nucleotide excision repair system displays a broad substrate specifity: NER recognizes all lesions that

generate bulky distortions in the conformation of the DNA double helix, e.g. UV-induced photolesions such as pyrimidine dimers, intrastrand-crosslinks and large chemical adducts in the DNA formed through exposure to genotoxic agents (Batty and Wood, 2000). Analyses of NER in *Bacillus subtilis* (Sancar, 1996), *Streptococcus pneumoniae* (Sicard *et al.*, 1992), *Mycoplasma genitalium* (Sancar, 1996) and *Deinococcus radiodurans* (Minton, 1994) indicate that the repair mechanism originally characterised in *E. coli* is highly conserved in prokaryotes. Interestingly, an additional function of mycobacterial UvrD besides NER has been proposed recently; UvrD was shown to physically and functionally interact with Ku, a protein participating in the NHEJ pathway of DNA double strand break repair (Sinha *et al.*, 2007).

To study whether the mycobacterial NER pathway contributes to compensating for the inherent lack of MMR, we here generated mutants deficient in the NER components *uvrD* and *uvrB*, as well as a mutant lacking both proteins. The NER mutants were assessed with regard to their susceptibilities to various DNA damaging agents. Using a plasmid-based assay we also studied a possible involvement of mycobacterial NER in DNA mismatch recognition. Our data demonstrate that NER in mycobacteria is functional and that UvrB and UvrD play important roles in DNA repair and mutation prevention. In addition, we propose an additional function of UvrD besides involvement in the NER system, namely restricting gene conversion.

Experimental procedures

Bacterial strains, media and culture conditions

E. coli strain XL1 blue (*Stratagene*) was used for cloning and propagation of plasmids. Bacterial cultures were grown in Luria Bertani (LB) medium at 37°C for 14-20 h. Cultures of *M. smegmatis* SMR5 *rrnB* (Sander *et al.*, 1996), a derivative of *M. smegmatis* mc^2 155, were grown in 7H9 medium at 37°C for 2-3 days. When appropriate, antibiotics were added at the following concentrations: ampicillin 100 µg/ml, hygromycin 100 µg/ml; kanamycin 50 µg/ml; streptomycin 100 µg/ml; apramycin 20 µg/ml; clarithromycin 50 µg/ml; rifampicin 175 µg/ml.

Generation of mycobacterial NER mutant strains

Allelic replacement techniques were used to generate *M. smegmatis* knockout mutants (Sander *et al.*, 1995). Using genomic DNA, 1-1.5 kbp fragments upstream (5') and downstream (3') of the target genes were amplified by PCR and cloned into pMCS5-*rpsL-hyg* for generation of unmarked knock-outs. The following primers were used: uvrD-1 / uvrD-2 for amplification of the *uvrD* upstream region; uvrD-3 / uvrD-4 for amplification of the *uvrD* downstream region; uvrB-1 / uvrB-2 for amplification of the *uvrB* upstream region and uvrB-3 / uvrB-4 for amplification of the *uvrB* downstream region. The resulting deletion alleles lack basepairs 219 to 2040 of the 2352 basepair *uvrD* (MSMEG5534) ORF and basepairs 262 to 1890 of the 2160 basepair *uvrB* (MSMEG3816) ORF. Primer sequences are given in table 1; cloning was verified by sequencing. Vectors were transformed into *M. smegmatis* SMR5 *rrnB*, a derivative of *M. smegmatis* mc^2 155 carrying a non-restrictive *rpsL* mutation conferring streptomycin resistance (Sander *et al.*, 1996). Transformants were plated on medium containing hygromycin to select for integration of the knockout vector by homologous recombination into the genome. Clones that had undergone a single cross-over event were detected by Southern blot analyses and subjected to counterselection. In brief, single cross-over clones were grown for three days in liquid broth and subsequently plated on selective agar containing streptomycin. Putative knock-out mutants were colony purified and investigated by Southern blot analyses for disruption of *uvrD* and *uvrB*, respectively. For generation of the *M. smegmatis uvrB / uvrD* double mutant, the *uvrB* knock-out vector was transformed into the *M. smegmatis uvrD* mutant and selection procedures were applied as detailed above. Gene disruption was confirmed by Southern blot analyses.

Determination of spontaneous mutation frequencies

Mutation frequencies were determined by fluctuation experiments. Briefly, at least six parallel cultures of each strain were grown until late log phase in 7H9 medium, with a viable cell number around 10^9/ml. Subsequently, cultures were diluted to $2x10^3$/ml and incubated for 2 days at 37°C. 100 µl of each culture were plated on freshly prepared 7H10 agar plates containing rifampicin and serial dilutions were plated on non-selective medium. After 4 days of incubation at 37°C, colony forming units (CFU) were determined. The number of CFUs obtained on agar plates containing rifampicin was divided by the number of CFUs obtained on nonselective medium. For calculation of mutation frequencies, the median of the ratio of cells that gained resistance against rifampicin was determined, as previously described (Lea and Coulson, 1949; Luria and Delbrück, 1943).

Determination of MIC and MBC

Minimal inhibitory concentrations (MIC) were determined by E-test (AB BIODISK) according to the manufacturer's instructions. Determination of MBCs was performed in a microtiter plate format as described previously (Pfister *et al.*, 2005). In brief, freshly grown cultures were diluted to an absorbance (A600) of 0.01 in 7H9 medium and incubated for 72 h at 37°C in the presence of 2-fold serial dilutions of rifampicin. The rifampicin stock solution (2 mg/ml) was made in DMSO. Aliquots from those wells that showed growth inhibition were plated on drug-free solid agar and incubated at 37°C for a further 72 h. The MBC is defined as the minimal drug concentration that kills > 99.9 % of the inoculum.

Survival after exposure to UV light

UV sensitivity of strains was determined by plating dilutions (in triplicate) on solid medium followed by irradiation of open plates in a Stratalinker 2400 (Stratagene, 254nm) at 0–160 mJ/cm^2 (approximately 0-5 seconds). CFUs were determined after 3-4 days incubation at 37°C and compared to the untreated control.

Survival after treatment with DNA damaging agents

25 ml cultures of each strain were grown in 7H9 medium until early mid-log phase and serial dilutions were plated on LB agar for CFU determination. For determination of susceptibility to $NaNO_2$, cultures were centrifuged and resuspended in 7H9 medium acidified with HCl (pH 5.4). Subsequently, cultures were split into six 3 ml aliquots. Freshly prepared compounds (*tert*-butylhydroperoxide 250 µM, $NaNO_2$ 3 mM, mitomycin C 0.02 µg/ml) were added to

three aliquots, the remaining aliquots served as untreated control. After incubation for 24 h at 37°C, serial dilutions were plated onto LB agar. CFUs were counted after 3 days growth at 37°C. Survival was calculated by the ratio of CFUs of the treated cultures compared to the CFUs of the untreated controls.

Determination of gene conversion frequencies

M. smegmatis strains were transformed with an integrative vector carrying various versions of mutated *rrnA* gene fragments. The approximately 1.0 kb partial 16S rRNA gene fragment with a mutated 1491 position covered 16S rRNA position 907 to ITS 1 position 2367, the partial 23S rRNA gene fragment with a mutated 2058 or 2059 position covered 23S rRNA position 1426 to 2624. Upon recombination with the chromosomal wildtype *rrnA* gene, the vector-encoded mutated *rrnA* gene fragment confers drug resistance (apramycin for the 16S rRNA 1491 mutation, clarithromycin for the 23S rRNA 2058/2059 mutation, vectors listed in table 2). Gene conversion frequencies were assessed by fluctuation experiments. Briefly, for each construct and strain, at least ten independent transformants were picked and grown. Aliquots of these precultures were taken to inoculate 4 ml 7H9 medium with approximately $5x10^5$ cells. After two days of growth, serial dilutions were plated on permissive medium and on selective medium containing apramycin or clarithromycin. The relative marker integration frequency was determined by the ratio of cells which gained antibiotic resistance compared to the number of cells obtained on permissive medium.

Results

To study the role of the NER pathway in mycobacteria, we generated *M. smegmatis* mutants deficient in *uvrB* and in *uvrD*. In addition, a double mutant deficient in both genes was constructed. Disruption was verified by Southern blot analysis (Figures 1a, 1b). For comparison, a *M. smegmatis recA* mutant (Frischkorn *et al.*, 1998) was included in the studies.

Growth characteristics. The *in vitro* growth characteristics of the *uvrB* and *uvrD* mutant strains in Middlebrook 7H9 medium were indistinguishable from that of the parental strain. Growth of the *uvrB* / *uvrD* double mutant was considerably impaired, with a twofold increase in generation time compared to the wildtype (Figure 2a).

Survival after treatment with DNA damaging agents. To investigate whether absence of NER in *M. smegmatis* affects DNA repair proficiency, mutants were challenged with DNA-damaging agents in survival assays. Different DNA damaging compounds were chosen to cover various types of DNA damage.

i) NER is responsible for the repair of bulky adducts resulting from UV-light exposure in *E. coli*, prompting us to study the impact of short wavelength UV light (UV-C) on survival of *M. smegmatis* NER mutants. The *uvrB* and *uvrD* mutant strains as well as the *recA* mutant strain were considerably more susceptible than the wildtype strain. Interestingly, combined deletion of *uvrB* and *uvrD* had an additive effect on survival following treatment with UV; the double mutant was unable to survive doses in excess of 80 mJ / cm^2 (Figure 2b). ii) *Tert*-butyl-hydroperoxide (TBH) generates ROI which can attack the bases and the sugar-phosphate backbone of the DNA, leading to strand breaks and base adducts (Halliwell and Aruoma, 1991). Treatment with TBH may also result in formation of alkoxyradicals and methylation reactions of the DNA, thus leading to alkylation damage (Hix *et al.*, 1995). The *M. smegmatis* NER mutants were significantly more susceptible towards TBH compared to the wildtype strain (Figure 2c; $p < 0.05$, student's t-test). The *uvrB*, *uvrD* and the *uvrB / uvrD* double mutant displayed similar susceptibilities (student's t-test: $p > 0.1$). TBH susceptibility of the *M. smegmatis recA* mutant was only slightly, yet significantly increased compared to the wildtype (42 % surviving cells *recA* mutant compared to 77% surviving cells wildtype; $p < 0.05$ student's t-test). These findings indicate that NER is more important than homologous recombinational repair in defence against ROI and alkylation damage in mycobacteria. iii) Acidified sodium nitrite generates RNI, leading to oxidative damage of the DNA as well as inducing nitration, nitrosation and deamination reactions (Burney *et al.*, 1999). The *M. smegmatis recA* mutant displayed a similar susceptibility to RNI as the wildtype (12% and 15

% surviving cells; $p > 0.08$ student's t-test, Figure 2d), while the *M. smegmatis* NER mutants displayed significantly increased susceptibilities to acidified sodium nitrate ($p < 0.05$, student's t-test). The *uvrD* and *uvrB / uvrD* mutants were considerably more susceptible than the *uvrB* mutant (Figure 2d), suggesting a further role of *uvrD* in the repair of RNI induced DNA damage. iv) Exposure to mitomycin C (MMC) results in alkylation and interstrand crosslinks (Iyer and Szybalski, 1963; Kumar *et al.*, 1992), which are mainly repaired by NER in *E. coli* (Van Houten *et al.*, 2005). The *M. smegmatis* NER mutants showed different susceptibilities towards treatment with MMC. While susceptibilities of the *recA* and the *uvrD* mutants were significantly increased compared to the wildtype, the *uvrB* and *uvrB / uvrD* mutants were highly susceptible towards MMC (Figure 2e).

Spontaneous mutation frequencies. Rifampicin has a single molecular target, the ß subunit of RNA polymerase encoded by *rpoB*. Various point mutations in *rpoB* confer high-level drug resistance. The frequency at which bacteria generate rifampicin-resistant mutants is widely used to assess the general spontaneous mutation frequency. The calculation is based on the model for the distribution of mutant clones originally described by Luria and Delbrück (1943) and extended by Lea and Coulson (1949). The wildtype strain displayed a spontaneous mutation frequency of 3.8×10^{-8} (Figure 3). In comparison, deletion of *uvrD* resulted in an approximately threefold increased mutation frequency (9.1×10^{-8}, $p < 0.05$, n= 6, students t-test) and deletion of *uvrB* resulted in a fivefold increased mutation frequency (1.9×10^{-7}, $p < 0.01$, n= 6, students t-test). Inactivation of *recA* did not affect the mutation frequency (1.7×10^{-8} $p > 0.05$, n= 6, students t-test). The mutation frequency of the double mutant could not be assessed with this approach, as the *uvrB / uvrD* mutant strain was unable to grow on plates containing 175 µg/ml rifampicin used for selection of spontaneous drug resistant mutants. Considering the general growth defect of this strain, we assumed that the *uvrB / uvrD* mutant is incapable of acquiring high-level resistance against rifampicin. Minimal inhibitory concentration (MIC) assays revealed that the *uvrB / uvrD* mutant displayed a decreased MIC - 4 µg/ml rifampicin compared to 32 µg/ml (wildtype, *uvrB* mutant strain) and 16 µg/ml (*uvrD* mutant strain); the minimal bactericidal concentration (MBC) for the *uvrB / uvrD* double mutant were found to be 130 µg /ml compared to 500 µg /ml for the wildtype strain and the *uvrB* and *uvrD* single mutant strains. Given the decreased MBC of rifampicin, we assessed the spontaneous mutation frequency of the *uvrB* / *uvrD* double mutant on plates containing 75 µg/ml rifampicin. Using this approach, the *uvrB* / *uvrD* double mutant strain was found to display an approximately 20 fold increased spontaneous mutation frequency (7.2×10^{-7}, $p < 0.01$, n= 6, students t-test) compared to the wildtype strain.

Involvement of NER in mismatch recognition. Base-base mispairing occurs in duplex DNA in the form of purine-purine (G/G, A/A, G/A), purine-pyrimidine (G/T, A/C) or pyrimidine-pyrimidine (C/C, T/T, T/C) mismatches. With the exception of C/C, all base-base mispairings are subject to correction by the mismatch repair system (Marra and Schär, 1999). As mycobacteria are devoid of this otherwise highly conserved repair system, the effects of MMR deficiency and a possible compensating role of NER in postreplicative repair were investigated. Towards this end, a plasmid transformation-based assay was applied (Pfister *et al.*, 2004). Following integration at the *attB* site of the mycobacterial genome, intramolecular recombination between the chromosomal *M. smegmatis rrnA* gene and a plasmid-borne *rrnA* gene fragment carrying a specific resistance mutation leads to antibiotic resistance. Intrachromosomal recombination occurs via an intermediate with a non-Watson-Crick base pairing. In repair proficient cells, the intermediate is recognized and rejected, resulting in decreased gene conversion frequencies. In case of repair deficiency, the intermediate will not be recognized as abnormal, leading to mutation of the chromosomal *rrnA* operon following replication and postmitotic segregation in 50 % of the progeny, with consequent gain in resistance to the relevant antibiotic. The assay provided the opportunity to analyse recognition of different mismatches, e.g. purine-pyrimidine (G/T, A/C) or pyrimidine-pyrimidine (C/C, C/T, T/C) mismatches. The *M. smegmatis* wildtype chromosome harbours a second *rrn* operon (named *rrnB*) in addition to the *rrnA* operon, which precludes proper measurement of gene conversion frequencies using the assay described above. Hence, the NER mutants were constructed in a Δ *rrnB* background (Sander *et al.*, 1996).

M. smegmatis rrnB NER mutant strains were transformed with an integrative vector carrying a mutated *rrnA* gene fragment. For each of the different *rrnA* fragments, specific gene conversion frequencies were obtained. Comparison of marker integration frequencies observed in the wildtype revealed that the highest gene conversion frequencies (in the range of 10^{-4} to 10^{-5}) were found with an A → C mutation leading to a C/T mismatch (Figure 4a). This mismatch was analysed making use of two different constructs with the A → C mutation located at 23S rRNA position 2058, or 2059, respectively. Deletion of *uvrD* resulted in significantly (approximately tenfold) increased gene conversion frequencies in comparison to the wildtype. In contrast, gene conversion frequencies in the *uvrB* mutant were virtually at wildtype control level (student's t-test, n=10, $p>0.05$). Interestingly, combined deletion of *uvrB* and *uvrD* resulted in significantly elevated marker integration frequencies (about 5 fold [A → C mutation located at 23S rRNA position 2058] or 20 fold [A → C mutation located at 23S rRNA position 2059]) compared to the frequencies obtained in the *uvrD* single mutant

(Figure 4a). Analysis of marker integration frequencies with an *rrnA* fragment carrying an A → G mutation which leads to a G/T mismatch, gave similar results (Figure 4a). Again, two different constructs carrying the mutation at 23S rRNA position 2058, or 2059, were used. In general, marker integration frequencies were lower with the A → G mutation than with the A → C mutation, yielding frequencies in the range of 10^{-5} and 10^{-6} in *M. smegmatis* wildtype. Gene conversion frequencies were considerably (about tenfold) elevated in the *uvrD* mutant, but only marginally (3 to 4 fold) increased in the *uvrB* mutant. A synergism of combined absence of *uvrB* and *uvrD* was observed, as the double mutant displayed marker integration frequencies about 20 fold higher than those of the *uvrD* single mutant (Figure 4a).

Recognition of A/C and C/C mismatches was determined with *rrnA* fragments carrying a G → A mutation or a G → C mutation at 16S rRNA position 1491. Gene conversion frequencies of the wildtype were in the range of 1 x 10^{-6}. Gene conversion frequencies of the *uvrB* mutant, and the *uvrD* mutant were only marginally (about fourfold) increased and no synergism of combined deletion of *uvrB* and *uvrD* was observed for A/C and C/C mismatch recognition (Figure 4b). Analysis of marker integration frequencies using an *rrnA* fragment carrying a G → T mutation provided insights into recognition of T/C mismatches. The *uvrD* mutant strain and the *uvrD / uvrB* mutant strain yielded more than tenfold increased gene conversion frequencies, whilst the *uvrB* mutant strain yielded fourfold increased frequencies (Figure 4b).

Discussion

In this study, *M. smegmatis* mutants deficient in NER were generated and characterised. NER mutants in *E. coli* are sensitive not only to irradiation with UV, but also to a wide range of chemical agents, including MMC (Friedberg, 2005). Consistently, we found that *M. smegmatis* NER mutants are sensitive to treatment with UV and MMC. Compared to the mycobacterial *uvrB* mutant, the *uvrD* mutant is extremely susceptible towards MMC, corroborating the previous observation that *E. coli uvrB* mutants were significantly more susceptible towards MMC than *E. coli uvrA*, *uvrC* and *uvrD* mutants (Vidal *et al.*, 2006). The *M. smegmatis* NER mutants, but not the *recA* mutant, are hypersensitive to TBH, indicating a role of mycobacterial NER in the repair of oxidative and alkylation damage. In contrast, repair of oxidative DNA damage is primarily accomplished by homologous recombinational repair and base excision repair (BER) in *E. coli* (Konola *et al.*, 2000). NER of *E. coli* is not involved in the repair of ROI induced lesions, as *E. coli uvrA, uvrB* and *uvrC* mutants show wild-type like sensitivity to hydrogen peroxide (Imlay and Linn, 1987). Alkylation damage is mainly repaired by methyltransferases and BER in *E. coli* (Nowosielska et al., 2006). NER is, if at all, only moderately capable of repairing these lesions (Samson *et al.*, 1988; Van Houten *et al.*, 2005). Our findings indicate that contrary to observations in *E. coli*, NER is more important than homologous recombinational repair in defence against ROI and alkylation damage in mycobacteria. Repair of RNI induced DNA damage in *E. coli* is mainly accomplished by repair mechanisms other than NER, primarily by BER and homologous recombinational repair (Spek *et al.*, 2001), although a role for UvrA in resistance to acidified sodium nitrite has been reported (Sidorkina *et al.*, 1997). The mycobacterial NER mutants are more susceptible than both the wildtype strain and the *recA* mutant to the RNI generating agent acidified sodium nitrite. The mild RNI / ROI phenotype of the *recA* mutant indicates a minor role for homologous recombinational repair in repairing ROI and RNI induced lesions. Taken together, our data provide evidence that NER in mycobacteria is functional and that UvrB and UvrD play important roles in DNA repair. This is corroborated by analyses of spontaneous mutation frequencies towards rifampicin resistance. *M. smegmatis uvrD* and *uvrB* mutants display elevated spontaneous mutation frequencies compared to the wildtype strain, indicating that disruption of NER has an impact on genomic integrity. Our findings suggest that the substrate spectrum of mycobacterial NER may differ from that of the described model species *E. coli*.

Combined deletion of *uvrB* and *uvrD* has not been analysed in other prokaryotes so far. Interestingly, loss of both proteins rendered cells more sensitive to UV than the *uvrB* and

uvrD single mutants, implying an additional role for one of these proteins besides NER. This assumption is corroborated by the observation that the *in vitro* growth rate of the double mutant was significantly reduced, whereas growth of the *uvrB* and *uvrD* single mutants was similar to that of the wildtype. We were unable to select for spontaneous resistant clones of the *uvrB / uvrD* mutant on medium containing high concentrations of rifampicin and found that the double mutant is significantly more susceptible to rifampicin as determined by MIC and MBC assays. Hence, the spontaneous mutation frequency towards rifampicin resistance was analysed at lower drug concentrations. The spontaneous mutation frequency of the double mutant was found to be considerably elevated compared to the *uvrB* and *uvrD* single mutants, indicating that overall repair capability is significantly impaired. However, a statistically significant additive effect of combined lack of *uvrB* and *uvrD* was not observed when the strains were assayed for their sensitivity to MMC, TBH and acidified sodium nitrate. Other repair pathways, like BER and recombinational repair, may have overlapping repair specificities for DNA damage generated by these compounds and thus may partially compensate for the lack of *uvrB* and *uvrD* under these conditions.

To investigate mismatch recognition in the absence of a MMR system, we made use of gene conversion assays. The assay applied in this study allows *M. smegmatis* cells to acquire antibiotic resistance upon recombination between the chromosomal *rrnA* operon and a plasmid borne partial gene fragment carrying a specific resistance mutation. This assay provided the opportunity to analyse recognition of different mismatches. Quantification of the relative marker integration frequencies revealed that marker integration depends on the specific mismatch, corresponding to observations formerly made in MMR proficient species. The efficiency of repair for different base-pairing errors varies and appears to depend on multiple factors such as the physical structure of the mismatch and the sequence context (Dohet *et al.*, 1985; Jones *et al.*, 1987; Marra and Schär, 1999). In *E. coli*, the correction of G/T, A/C, G/G and A/A mismatches by the MMR system is highly effective, while elimination of C/T, T/T and A/G mismatches occurs with decreased efficiency. For C/C mispairing, no correction activity can be observed (Dohet *et al.*, 1985; Kramer *et al.*, 1984; Parker and Marinus, 1992). We found that the *M. smegmatis uvrB* mutant displayed elevated gene conversion frequencies for most of the mismatches analysed. As gene conversion is mediated by recombination and occurs as consequence of absent mismatch recognition, our results indicate that mycobacterial NER may be involved in limiting recombination and act in the repair of DNA mismatches. A complementary function of NER in recognition of mismatches is corroborated by studies in *Schizosaccharomyces pombe*. NER of *S. pombe* is

capable of recognizing and repairing C/C mismatches, which are not substrates of MMR. In the absence of MMR, other mismatches are processed by NER as efficiently as C/C mismatches (Fleck *et al.*, 1999; Kunz and Fleck, 2001). Marker integration frequencies obtained in the *M. smegmatis uvrD* mutant are significantly increased compared to those obtained in the *uvrB* mutant. These findings imply that UvrD has a further role besides involvement in the NER pathway. The involvement of mycobacterial UvrD in mismatch recognition and limiting recombination is supported by previous findings. Analysis of *E. coli* UvrD function (Washburn and Kushner, 1991) and more recent studies in *H. pylori* (Kang and Blaser, 2006) revealed that deficiency in UvrD results in a hyperrecombinogenic phenotype. These observations can be explained by two different models. The first model refers to the assumption that NER or MMR remains incomplete as a consequence of UvrD deficiency (Arthur and Lloyd, 1980). Thus, nicks in the DNA occur more frequently. Replication forks will be stalled at such nicks, subsequently leading to the formation of double strand breaks which will be repaired by recombinational repair (Arthur and Lloyd, 1980). In another model it was proposed that UvrD participates in degradation of toxic recombination intermediates by actively removing RecA proteins from DNA (Centore and Sandler, 2007; Lestini and Michel, 2007). This model is based on observations made in yeast and *E. coli*. UvrD of *E. coli* (Centore and Sandler, 2007; Flores *et al.*, 2004; Flores *et al.*, 2005; Petit and Ehrlich, 2002; Veaute *et al.*, 2005) as well as the helicase Srs2 of yeast (Veaute *et al.*, 2003) were found to act as anti-recombinases to prevent potentially deleterious recombination events.

The *uvrB* / *uvrD* double mutant displayed significantly higher gene conversion frequencies than the *uvrB* and *uvrD* single mutants. Hence, combined loss of UvrB and UvrD function results in a synergistic (i.e. greater than additive) effect in mycobacteria. In the absence of NER and the anti-recombination function of UvrD, the bacteria are devoid of two important mutation avoidance mechanisms, resulting in high gene conversion frequencies. The synergism of UvrB and UvrD function in gene conversion prevention was not observed for recognition of A/C or C/C mismatches. The single mutant and the double mutant strains displayed similar marker integration frequencies, which were only marginally elevated compared to the wildtype strain. In the wildtype strain, A/C and C/C gene conversion frequencies are lower than those obtained with the other mismatches. These findings can be explained by the topological properties of the respective mismatches. A/C and C/C mismatches are the least stable mismatches and lead to significant distortions of the DNA double helix. Hence, due to the unstable nature of the mismatch, the mismatch is rejected irrespective of repair, which accounts for the low marker integration frequencies obtained in

the assay. Consequently, only a minor proportion of mismatch intermediate rejection is caused by the anti-recombination function of UvrD.

Taken together, our observations provide evidence that despite the lack of MMR, mycobacteria possess efficient DNA repair mechanisms. We found that mycobacterial NER is capable of repairing a wide range of DNA damage. Furthermore, our data suggest a complementary function of NER in the recognition of DNA mismatches. The helicase component UvrD was shown to limit marker integration frequencies. We propose that repair of DNA mismatches is accomplished differently in mycobacteria than in the described model species. Further investigations will provide insights into new mechanisms that govern genome stability in mycobacteria.

Acknowledgements

This work was supported in part by grants from the University of Zurich (to B. S.), the Swiss National Science Foundation (to P. S. and E.C.B., BO-3200-68488) and the European Community (to B.S., E.C.B. and E.O.D.; CSI_LTB, LSHP-CT-2007-037235). E.O.D. was supported by the Medical Research Council (UK).

References

Arthur, H.M., and Lloyd, R.G. (1980) Hyper-recombination in *uvrD* mutants of *Escherichia coli* K-12. *Mol Gen Genet* **180**: 185-191.

Batty, D.P., and Wood, R.D. (2000) Damage recognition in nucleotide excision repair of DNA. *Gene* **241**: 193-204.

Burney, S., Caulfield, J.L., Niles, J.C., Wishnok, J.S., and Tannenbaum, S.R. (1999) The chemistry of DNA damage from nitric oxide and peroxynitrite. *Mutat Res* **424**: 37-49.

Centore, R.C., and Sandler, S.J. (2007) UvrD limits the number and intensities of RecA-green fluorescent protein structures in *Escherichia coli* K-12. *J Bacteriol* **189**: 2915-2920.

Chan, J., Xing, Y., Magliozzo, R.S., and Bloom, B.R. (1992) Killing of virulent *Mycobacterium tuberculosis* by reactive nitrogen intermediates produced by activated murine macrophages. *J Exp Med* **175**: 1111-1122.

Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry, C.E., 3rd, Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., and Barrell, B.G. (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.

Cole, S.T., Eiglmeier, K., Parkhill, J., James, K.D., Thomson, N.R., Wheeler, P.R., Honore, N., Garnier, T., Churcher, C., Harris, D., Mungall, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R.M., Devlin, K., Duthoy, S., Feltwell, T., Fraser, A., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Lacroix, C., Maclean, J., Moule, S., Murphy, L., Oliver, K., Quail, M.A., Rajandream, M.-A., Rutherford, K.M., Rutter, S., Seeger, K., Simon, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Taylor, K., Whitehead, S., Woodward, J.R., and Barrell, B.G. (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**: 1007-1011.

Curti, E., Smerdon, S.J., and Davis, E.O. (2007) Characterization of the helicase activity and substrate specificity of *Mycobacterium tuberculosis UvrD*. *J Bacteriol* **189**: 1542-1555.

Ding, A., Nathan, C., and Stuehr, D. (1988) Release of reactive nitrogen intermediates and reactive oxygen intermediates from mouse peritoneal macrophages. Comparison of activating cytokines and evidence for independent production. *J Immunol* **141**: 2407-2412.

Dohet, C., Wagner, R., and Radman, M. (1985) Repair of defined single base-pair mismatches in *Escherichia coli*. *Proc Natl Acad Sci U S A* **82**: 503-505.

Durbach, S.I., Springer, B., Machowski, E.E., North, R.J., Papavinasasundaram, K.G., Colston, M.J., Böttger, E.C., and Mizrahi, V. (2003) DNA alkylation damage as a sensor of nitrosative stress in *Mycobacterium tuberculosis*. *Infect Immun* **71**: 997-1000.

Fleck, O., Lehmann, E., Schar, P., and Kohli, J. (1999) Involvement of nucleotide-excision repair in *msh2 pms1*-independent mismatch repair. *Nat Genet* **21**: 314-317.

Flores, M.J., Bidnenko, V., and Michel, B. (2004) The DNA repair helicase UvrD is essential for replication fork reversal in replication mutants. *EMBO Rep* **5**: 983-988.

Flores, M.J., Sanchez, N., and Michel, B. (2005) A fork-clearing role for UvrD. *Mol Microbiol* **57**: 1664-1675.

Friedberg, E.C., Graham C. Walker, Wolfram Siede, Richard D. Wood, Roger A. Schultz, and Tom Ellenberger (2005) *DNA Repair and Mutagenesis.* Washington DC: American Society of Microbiology Press.

Frischkorn, K., Sander, P., Scholz, M., Teschner, K., Prammananan, T., and Böttger, E.C. (1998) Investigation of mycobacterial *recA* function: protein introns in the RecA of pathogenic mycobacteria do not affect competency for homologous recombination. *Mol Microbiol* **29**: 1203-1214.

Garnier, T., Eiglmeier, K., Camus, J.-C., Medina, N., Mansoor, H., Pryor, M., Duthoy, S., Grondin, S., Lacroix, C., Monsempe, C., Simon, S., Harris, B., Atkin, R., Doggett, J., Mayes, R., Keating, L., Wheeler, P.R., Parkhill, J., Barrell, B.G., Cole, S.T., Gordon, S.V., and Hewinson, R.G. (2003) The complete genome sequence of *Mycobacterium bovis*. *Proc Natl Acad Sci U S A* **100**: 7877-7882.

Halliwell, B., and Aruoma, O.I. (1991) DNA damage by oxygen-derived species. Its mechanism and measurement in mammalian systems. *FEBS Lett* **281**: 9-19.

Hix, S., Morais Mda, S., and Augusto, O. (1995) DNA methylation by tert-butyl hydroperoxide-iron (II). *Free Radic Biol Med* **19**: 293-301.

Imlay, J.A., and Linn, S. (1987) Mutagenesis and stress responses induced in *Escherichia coli* by hydrogen peroxide. *J Bacteriol* **169**: 2967-2976.

Iyer, V.N., and Szybalski, W. (1963) A Molecular Mechanism of Mitomycin Action: Linking of Complementary DNA Strands. *Proc Natl Acad Sci U S A* **50**: 355-362.

Jones, M., Wagner, R., and Radman, M. (1987) Repair of a mismatch is influenced by the base composition of the surrounding nucleotide sequence. *Genetics* **115**: 605-610.

Kang, J., and Blaser, M.J. (2006) UvrD helicase suppresses recombination and DNA damage-induced deletions. *J Bacteriol* **188**: 5450-5459.

Kaufmann, S.H. (2001) How can immunology contribute to the control of tuberculosis? *Nat Rev Immunol* **1**: 20-30.

Konola, J.T., Sargent, K.E., and Gow, J.B. (2000) Efficient repair of hydrogen peroxide-induced DNA damage by *Escherichia coli* requires SOS induction of RecA and RuvA proteins. *Mutat Res* **459**: 187-194.

Kramer, B., Kramer, W., and Fritz, H.J. (1984) Different base/base mismatches are corrected with different efficiencies by the methyl-directed DNA mismatch-repair system of *E. coli*. *Cell* **38**: 879-887.

Kumar, S., Lipman, R., and Tomasz, M. (1992) Recognition of specific DNA sequences by mitomycin C for alkylation. *Biochemistry* **31**: 1399-1407.

Kunkel, T.A., and Erie, D.A. (2005) DNA mismatch repair. *Annu Rev Biochem* **74**: 681-710.

Kunz, C., and Fleck, O. (2001) Role of the DNA repair nucleases Rad13, Rad2 and Uve1 of *Schizosaccharomyces pombe* in mismatch correction. *J Mol Biol* **313**: 241-253.

Lea, D.E., and Coulson, C.A. (1949) The distribution of the numbers of mutants in bacterial populations. *J Gen* **49**: 264-285.

Lestini, R., and Michel, B. (2007) UvrD controls the access of recombination proteins to blocked replication forks. *Embo J*.

Luria, S.E., and Delbrück, M. (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**: 491-511.

Marra, G., and Schär, P. (1999) Recognition of DNA alterations by the mismatch repair system. *Biochem J* **338 (Pt 1)**: 1-13.

Matic, I., Rayssiguier, C., and Radman, M. (1995) Interspecies gene exchange in bacteria: The role of SOS and mismatch repair systems in evolution of species. *Cell* **80**: 507-515.

Minton, K.W. (1994) DNA repair in the extremely radioresistant bacterium *Deinococcus radiodurans*. *Mol Microbiol* **13**: 9-15.

Mizrahi, V., and Andersen, S.J. (1998) DNA repair in *Mycobacterium tuberculosis*. What have we learnt from the genome sequence? *Mol Microbiol* **29**: 1331-1339.

Nathan, C., and Shiloh, M.U. (2000) Reactive oxygen and nitrogen intermediates in the relationship between mammalian hosts and microbial pathogens. *Proc Natl Acad Sci U S A* **97**: 8841-8848.

Nowosielska, A., Smith, S.A., Engelward, B.P., and Marinus, M.G. (2006) Homologous recombination prevents methylation-induced toxicity in *Escherichia coli*. *Nucleic Acids Res* **34**: 2258-2268.

Parker, B.O., and Marinus, M.G. (1992) Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. *Proc Natl Acad Sci U S A* **89**: 1730-1734.

Petit, M.A., and Ehrlich, D. (2002) Essential bacterial helicases that counteract the toxicity of recombination proteins. *EMBO J* **21**: 3137-3147.

Pfister, P., Jenni, S., Poehlsgaard, J., Thomas, A., Douthwaite, S., Ban, N., and Böttger, E.C. (2004) The structural basis of macrolide-ribosome binding assessed using mutagenesis of 23 S rRNA positions 2058 and 2059. *J Mol Biol* **342**: 1569-1581.

Pfister, P., Hobbie, S., Brüll, C., Corti, N., Vasella, A., Westhof, E., and Böttger, E.C. (2005) Mutagenesis of 16 S *rRNA* C1409-G1491 base-pair differentiates between 6'OH and 6'NH3+ aminoglycosides. *J Mol Biol* **346**: 467-475.

Rayssiguier, C., Thaler, D.S., and Radman, M. (1989) The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. *Nature* **342**: 396-401.

Samson, L., Thomale, J., and Rajewsky, M.F. (1988) Alternative pathways for the *in vivo* repair of O6-alkylguanine and O4-alkylthymine in *Escherichia coli*: the adaptive response and nucleotide excision repair. *Embo J* **7**: 2261-2267.

Sancar, A. (1996) DNA Excision Repair. *Annu Rev Biochem* **65**: 43-81.

Sander, P., Meier, A., and Böttger, E.C. (1995) *rpsL*+: a dominant selectable marker for gene replacement in mycobacteria. *Mol Microbiol* **16**: 991-1000.

Sander, P., Prammananan, T., and Böttger, E.C. (1996) Introducing mutations into a chromosomal *rRNA* gene using a genetically modified eubacterial host with a single *rRNA* operon. *Mol Microbiol* **22**: 841-848.

Sander, P., Papavinasasundaram, K.G., Dick, T., Stavropoulos, E., Ellrott, K., Springer, B., Colston, M.J., and Böttger, E.C. (2001) *Mycobacterium bovis* BCG *recA* deletion mutant shows increased susceptibility to DNA-damaging agents but wild-type survival in a mouse infection model. *Infect Immun* **69**: 3562–3568.

Schofield, M.J., and Hsieh, P. (2003) DNA mismatch repair: molecular mechanisms and biological function. *Annu Rev Microbiol* **57**: 579-608.

Sicard, N., Oreglia, J., and Estevenon, A.M. (1992) Structure of the gene complementing uvr-402 in *Streptococcus pneumoniae*: homology with *Escherichia coli uvrB* and the homologous gene in *Micrococcus luteus*. *J Bacteriol* **174**: 2412-2415.

Sidorkina, O., Saparbaev, M., and Laval, J. (1997) Effects of nitrous acid treatment on the survival and mutagenesis of *Escherichia coli* cells lacking base excision repair (hypoxanthine-DNA glycosylase-ALK A protein) and/or nucleotide excision repair. *Mutagenesis* **12**: 23-28.

Sinha, K.M., Stephanou, N.C., Gao, F., Glickman, M.S., and Shuman, S. (2007) Mycobacterial UvrD1 is a Ku-dependent DNA helicase that plays a role in multiple DNA repair events, including double-strand break repair. *J Biol Chem* **282**: 15114-15125.

Spek, E.J., Wright, T.L., Stitt, M.S., Taghizadeh, N.R., Tannenbaum, S.R., Marinus, M.G., and Engelward, B.P. (2001) Recombinational repair is critical for survival of *Escherichia coli* exposed to nitric oxide. *J Bacteriol* **183**: 131-138.

Springer, B., Master, S., Sander, P., Zahrt, T., McFalone, M., Song, J., Papavinasasundaram, K.G., Colston, M.J., Böttger, E.C., and Deretic, V. (2001) Silencing of oxidative stress response in *Mycobacterium tuberculosis*: Expression patterns of *ahpC* in virulent and avirulent strains and effect of *ahpC* inactivation. *Infect Immun* **69**: 5967-5973.

Springer, B., Sander, P., Sedlacek, L., Hardt, W.-D., Mizrahi, V., Schar, P., and Böttger, E.C. (2004) Lack of mismatch correction facilitates genome evolution in mycobacteria. *Mol Microbiol* **53**: 1601-1609.

Van Houten, B., Croteau, D.L., DellaVecchia, M.J., Wang, H., and Kisker, C. (2005) 'Close-fitting sleeves': DNA damage recognition by the UvrABC nuclease system. *Mutat Res* **577**: 92-117.

Veaute, X., Jeusset, J., Soustelle, C., Kowalczykowski, S.C., Le Cam, E., and Fabre, F. (2003) The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**: 309-312.

Veaute, X., Delmas, S., Selva, M., Jeusset, J., Le Cam, E., Matic, I., Fabre, F., and Petit, M.A. (2005) UvrD helicase, unlike Rep helicase, dismantles RecA nucleoprotein filaments in *Escherichia coli*. *EMBO J* **24**: 180-189.

Vidal, L.S., Santos, L.B., Lage, C., and Leitao, A.C. (2006) Enhanced sensitivity of *Escherichia coli uvrB* mutants to mitomycin C points to a UV-C distinct repair for DNA adducts. *Chem Res Toxicol* **19**: 1351-1356.

Washburn, B.K., and Kushner, S.R. (1991) Construction and analysis of deletions in the structural gene (*uvrD*) for DNA helicase II of *Escherichia coli*. *J Bacteriol* **173**: 2569-2575.

Worth, L., Jr., Clark, S., Radman, M., and Modrich, P. (1994) Mismatch repair proteins MutS and MutL inhibit RecA-catalyzed strand transfer between diverged DNAs. *Proc Natl Acad Sci U S A* **91**: 3238-3241.

Table 1: Primers used for gene cloning

Primer name	Sequence
uvrD-1	5′-GGAATTCCATATGGTGAGGACGCCTACGAC-3′
uvrD-2	5′-GGAAGATCTGGTGAACGTGATGGCCAG-3′
uvrD-3	5′-GGAAGATCTGCAGGAACTCATCGACTGGCG-3′
uvrD-4	5′-TGCATGCATGATCGCGTCGGGCACCTTC-3′
uvrB-1	5′-GGAATTCCATATGGGCCGAGTACGGCCAGTC-3′
uvrB-2	5′-GGAAGATCTGCGCCATCACGAGCGTG-3′
uvrB-3	5′-GGAAGATCTGAGTCGGTCGAGATCGGTGG-3′
uvrB-4	5′-TGCATGCATCGTCATGTGCGCCAGCCGC-3′

Table 2: Recombination substrates containing a mutated *rrnA* fragment. All vectors harbour a mutated *rrnA* fragment which confers antibiotic resistance (corresponding antibiotic denoted in parentheses) upon recombination with the chromosomal wildtype *rrnA* gene. Numbers indicate the *rrnA* position of the mutated nucleotide.

pMV361ΔKan-Gm-rRNA A 2059 C (Clarithromycin)	(Pfister *et al.*, 2004)
pMV361ΔKan-Gm-rRNA A 2059 G (Clarithromycin)	(Pfister *et al.*, 2004)
pMV361ΔKan-Gm-rRNA A 2058 C (Clarithromycin)	(Pfister *et al.*, 2004)
pMV361ΔKan-Gm-rRNA A 2058 G (Clarithromycin)	(Pfister *et al.*, 2004)
pMV361ΔKan-Gm-rRNA G 1491 A (Apramycin)	(Pfister *et al.*, 2005)
pMV361ΔKan-Gm-rRNA G 1491 C (Apramycin)	(Pfister *et al.*, 2005)
pMV361ΔKan-Gm-rRNA G 1491 T (Apramycin)	(Pfister *et al.*, 2005)

Figure legends

Figure 1: Generation of *M. smegmatis* NER mutants

A Disruption of *M. smegmatis uvrD*. Left: Southern blot analysis. Genomic DNA from *M. smegmatis* wildtype (lane 1), *uvrD* mutant (lane 2) and *uvrD* single cross-over mutant (lane 3) was digested with *Bst*XI and probed with a 653bp *Mlu*I/*Bam*HI DNA fragment containing 5' flanking sequences of the *uvrD* gene. The presence of a single 10.7 kbp fragment instead of a 2.0 kbp fragment as seen in the parental strain demonstrates successful deletion of *uvrD* coding sequences. Right: Schematic illustration of the *uvrD* locus and Southern blot analysis. Shown are the genomic organization of the wild type (wt), the knockout vector that contains the *uvrD* deletion allele (vector), the single cross-over genotype (sco), and the mutated genomic *uvrD* region in the knockout mutant (ko). Fragments detected by the probe specific for the 5' flanking region are indicated.

B Disruption of *M. smegmatis uvrB.* Left: Southern blot analysis. Genomic DNA from *M. smegmatis uvrB* mutant (lane 1), *uvrB* single cross-over mutant (lane 2) and wildtype (lane 3) was digested with *Nco*I and probed with a 600bp *Nde*I/*Bsm*I *uvrB* gene fragment. The presence of a single 3.9 kbp fragment instead of a 5.6 kbp fragment as seen in the parental strain demonstrates successful deletion of *uvrB* coding sequences. Right: Schematic illustration of the *uvrB* locus and Southern blot analysis. Shown are the genomic organization of the wild type (wt), the knockout vector that contains the *uvrB* deletion allele (vector), the single cross-over genotype (sco), and the mutated genomic *uvrB* region in the knockout mutant (ko). Fragments detected by the probe specific for the *uvrB* gene are indicated.

Figure 2: *In vitro* growth analysis and survival of *M. smegmatis* strains following DNA damage

A *In vitro* growth analysis of *M. smegmatis* strains. Strains were grown in 7H9 broth in shaking cultures for 2 days at 37°C. Doubling times: wildtype 3.7 h +/- 0.1, *uvrD* mutant 3.5 h +/- 0.2, *uvrB* mutant 3.4 h +/- 0.2, *uvrB* / *uvrD* mutant 7.6 h +/- 0.4, recA mutant 3.9 h +/- 0.2.

B Survival following irradiation with UV-C. Survival was calculated from the ratio of cells that survived treatment with UV-C in comparison to the untreated control. Shown are the mean values of one representative experiment performed with three independent cultures each. Standard deviations of the means are given in the error bars. Ratio of treated to untreated cells is plotted in logarithmic scale. The *uvrB* / *D* mutant was unable to survive doses higher than 80 mJ / cm^2.

C-E Survival after treatment with different DNA damaging agents. C *tert*-butyl-hydroperoxide (250 µM) **D** acidified $NaNO_2$ (pH 5.4, 3mM); **E** Mitomycin C (0.02 µg/ml). Survival was calculated from the ratio of cells that survived treatment with the DNA damaging agent in comparison to the untreated control. Shown are the mean values of one representative experiment performed with three independent cultures each. Strains were incubated with the DNA damaging agent for 24 h. Standard deviations from the mean are given in the error bars. Abbreviations: wt = *M. smegmatis* wildtype strain; *uvrB*, *uvrD*, *uvrB*/*D* = *M. smegmatis* NER mutant strains, *recA* = *M. smegmatis recA* mutant strain.

Figure 3: Frequencies of mutation to rifampicin resistance

Shown are the median values (horizontal lines), 95% confidence interval (rectangular boxes) and outliers (vertical lines) of mutation frequencies calculated from the ratio of cells that gained rifampicin resistance in six independent cultures of each strain. * Significant increase of mutation frequency compared to the wildtype (student's t-test $p<0.05$). Abbreviations: wt = *M. smegmatis* wildtype strain; *uvrB*, *uvrD*, *uvrB*/*uvrD* = *M. smegmatis* NER mutant strains, *recA* = *M. smegmatis recA* mutant strain.

Figure 4 Relative gene conversion frequencies obtained with *rrnA* fragments resulting in C/T or G/ T mismatches (A) and *rrnA* fragments resulting in A/C, C/C or T/C mismatches (B)

Shown are the median values (horizontal lines), 95% confidence intervals (rectangular boxes) and outliers (vertical lines) of gene conversion frequencies calculated from the number of cells that gained apramycin or clarithromycin resistance in ten independent cultures of each strain. * significant increase of gene conversion frequencies compared to the wildtype (student's t-test, $p<0.05$). Abbreviations: wt = *M. smegmatis* wildtype strain; *uvrB*, *uvrD*, *uvrB*/*uvrD* = *M. smegmatis* NER mutant strains.

Figure 1

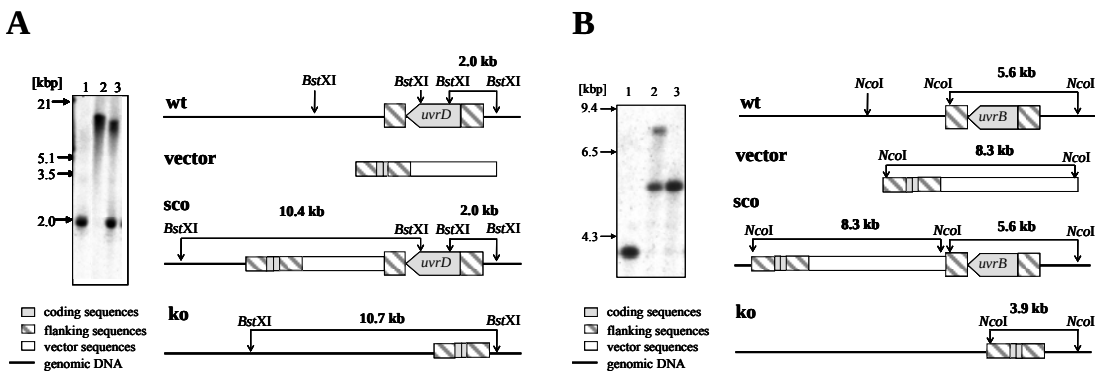

A
[kbp]
1 2 3
21
5.1
3.5
2.0
BstXI
2.0 kb
wt
uvrD
vector
sco
10.4 kb
ko
10.7 kb
coding sequences
flanking sequences
vector sequences
genomic DNA
B
9.4
6.5
4.3
NcoI
5.6 kb
uvrB
8.3 kb
3.9 kb

Figure 2

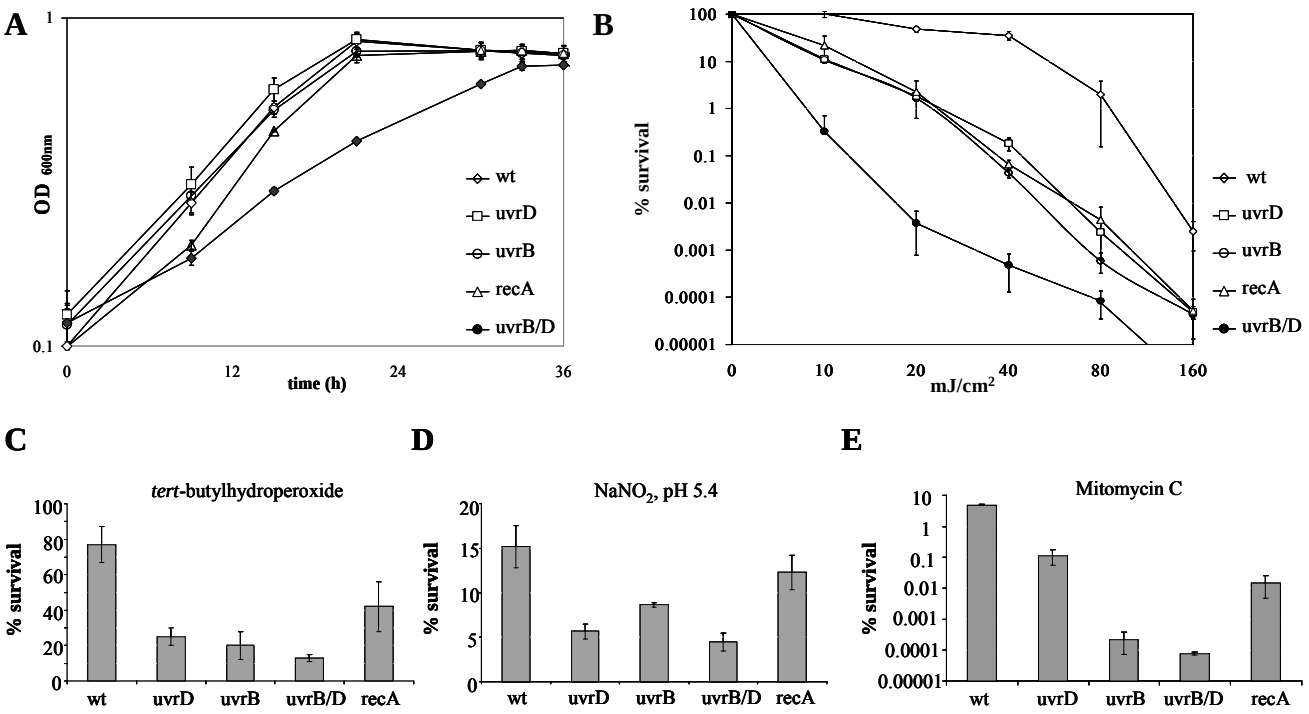

A
OD 600nm
time (h)
wt
uvrD
uvrB
recA
uvrB/D
B
% survival
mJ/cm2
C
tert-butylhydroperoxide
% survival
D
NaNO2, pH 5.4
% survival
E
Mitomycin C
% survival

Figure 3

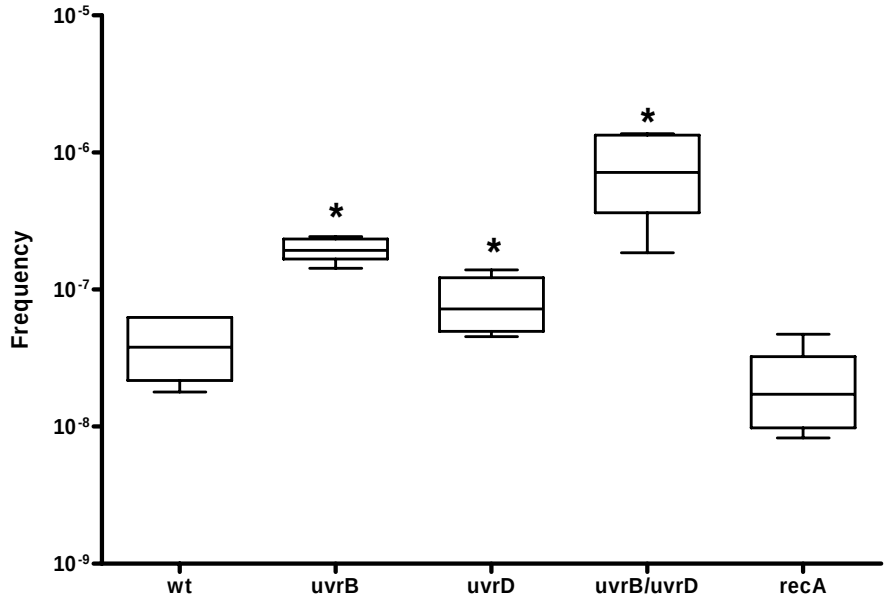
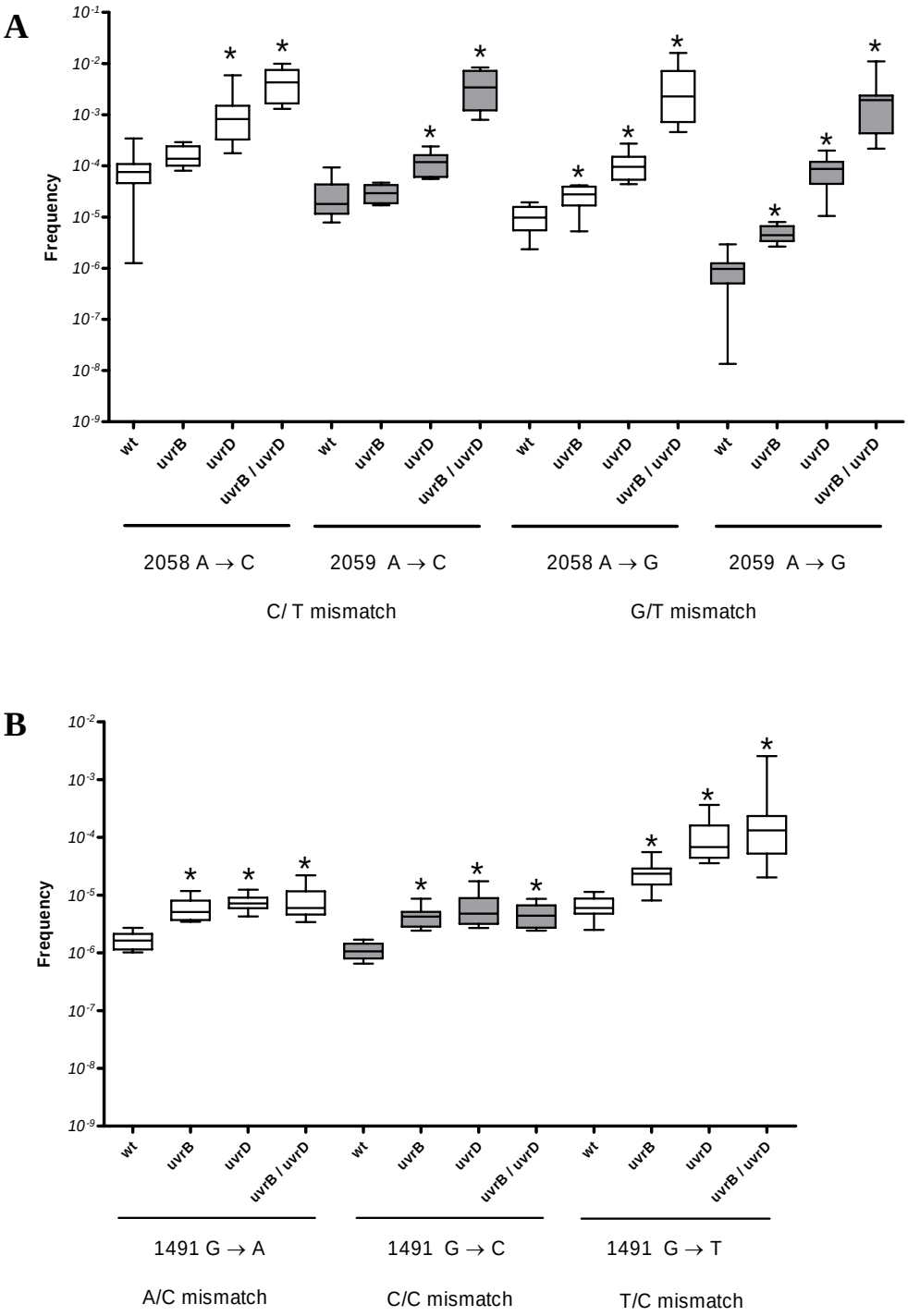


Figure 4


A
Frequency
wt
uvrB
uvrD
uvrB / uvrD
2058 A → C
2059 A → C
2058 A → G
2059 A → G
C/ T mismatch
G/T mismatch
B
1491 G → A
1491 G → C
1491 G → T
A/C mismatch
C/C mismatch
T/C mismatch

JOURNAL OF BACTERIOLOGY, Jan. 2008, p. 452–456 Vol. 190, No. 1
0021-9193/08/$08.00+0 doi:10.1128/JB.01315-07

A Mycobacterial *smc* Null Mutant Is Proficient in DNA Repair and Long-Term Survival[▽]

Carolin Güthlein,* Roger M. Wanner, Peter Sander, Erik C. Böttger, and Burkhard Springer†

Institut für Medizinische Mikrobiologie, Universität Zürich, Zurich, Switzerland, and Nationales Zentrum für Mykobakterien, Gloriastr. 30, CH-8006 Zurich, Switzerland[1]

Received 14 August 2007/Accepted 21 October 2007

SMC (structural maintenance of chromosomes) proteins play fundamental roles in various aspects of chromosome organization and dynamics, including repair of DNA damage. Mutant strains of *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* defective in SMC were constructed. Surprisingly, inactivation of *smc* did not result in recognizable phenotypes in hallmark assays characteristic for the function of these genes. This is in contrast to data for *smc* null mutants in other species.

Most bacterial cells contain a single circular chromosome that is folded and compacted into a structure called the nucleoid (24). Diverse DNA binding proteins are associated with the nucleoid and take part in chromosome organization to assist in a variety of complex processes such as replication, recombination, repair, modification, and transcription (8). Among these proteins are the SMC (structural maintenance of chromosomes) proteins that are conserved from prokaryotes to eukaryotes (4, 5, 28). The SMC family proteins are large proteins in the range between 110 and 170 kDa and share common principles in domain organization: globular N- and C-terminal domains which are connected by two long coiled-coil domains, separated by a globular hinge domain of approximately 150 amino acids in length (12).

Eukaryotes possess at least six distinct SMCs that participate in chromosome condensation, sister chromatid cohesion, DNA repair, and gene dosage compensation (16, 19). Pairs of SMC proteins form antiparallel heterodimers which assemble with accessory proteins into complexes, such as the condensin complex (SMC2 and SMC4), which condenses the chromosomes during mitosis, and the cohesin complex (SMC1 and SMC3), which mediates sister chromatid cohesion. Another complex composed of SMC5 and SMC6 is involved in the cellular response to DNA damage. Genetic and biochemical evidence suggests that cohesin and the SMC5/6 complex cooperate in the recombinational repair of double-strand breaks (17). The cohesin complex assists in homologous recombination by holding sister chromatids together in the vicinity of the double-strand breaks (26). The function of the SMC5/6 complex was first recognized by the increased sensitivity of *Schizosaccharomyces pombe* SMC6 mutants to UV light and ionizing radiation (21). Recent investigations of the SMC5/6 complex in yeasts, plants, and human cells have shown that deletion of components of this complex leads to an increased sensitivity towards DNA-damaging agents (22).

In contrast to the multiple SMC proteins present in eukaryotes, most bacteria possess only a single SMC (28). This single SMC protein is proposed to carry out several of the functions divided between specialized SMC complexes in eukaryotic cells. Bacterial SMC proteins form antiparallel homodimers through interactions in the hinge region, resulting in symmetric molecules where both ends contain an ATPase and a DNA binding domain (10). The gammaproteobacteria (such as *Escherichia coli*) possess the MukB protein instead of SMC. MukB is distantly related to SMC with structural and functional similarity (20). SMC from *Bacillus subtilis* or *Caulobacter crescentus* functionally resembles eukaryotic SMCs. The proteins are not essential, but null mutants of SMC generate anucleate cells ("titan cells") and display aberrant nucleoid formation (decondensed chromosomes) (2, 15, 31). In addition, deletion of SMC results in increased susceptibility to DNA-damaging agents (7).

Little is known about chromosome stability, organization, and partitioning in mycobacteria. *Mycobacterium tuberculosis* persists for decades in its host during latent infection (30). It is thus of interest to investigate the mechanisms which contribute to the capability of the bacterium to stabilize its genome during persistence and ensure proper cell division following reactivation. Bioinformatic analyses indicated that all mycobacterial genomes encode a single SMC homologue. A homologous open reading frame (ORF) is also present in *Mycobacterium leprae*, a bacterium that has lost various gene functions during reductive genomic evolution (6).

Mycobacterium smegmatis smc (MSMEG2422) encodes an 1,195-amino-acid polypeptide with a calculated molecular mass of 129.7 kDa. The protein shares all structural motifs conserved in members of the SMC family as determined by Clustal W multiple sequence alignments and InterPro Scan protein motif searches. Homologous sequences are restricted to the head, hinge, and tail domains. The three domains are connected by two long heptad repeat regions predicted to form coiled coils. Overall, *M. smegmatis* SMC shares 74% identity with the 1,205-amino-acid SMC sequence of *M. tuberculosis* and 55% identity with the 1,186-amino-acid SMC sequence of *B. subtilis*. Eukaryotic SMC proteins functionally interact with

* Corresponding author. Mailing address: Institut für Medizinische Mikrobiologie, Universität Zürich, Gloriastrasse 30/32, 8006 Zurich, Switzerland. Phone: 41 44 63 42 693. Fax: 41 44 634 49 06. E-mail: cguethl@immv.uzh.ch.

† Present address: Institute for Medical Microbiology and Hygiene, Austrian Agency for Health and Food Safety, Beethovenstrasse 6, 8010 Graz, Austria.

▽ Published ahead of print on 2 November 2007.

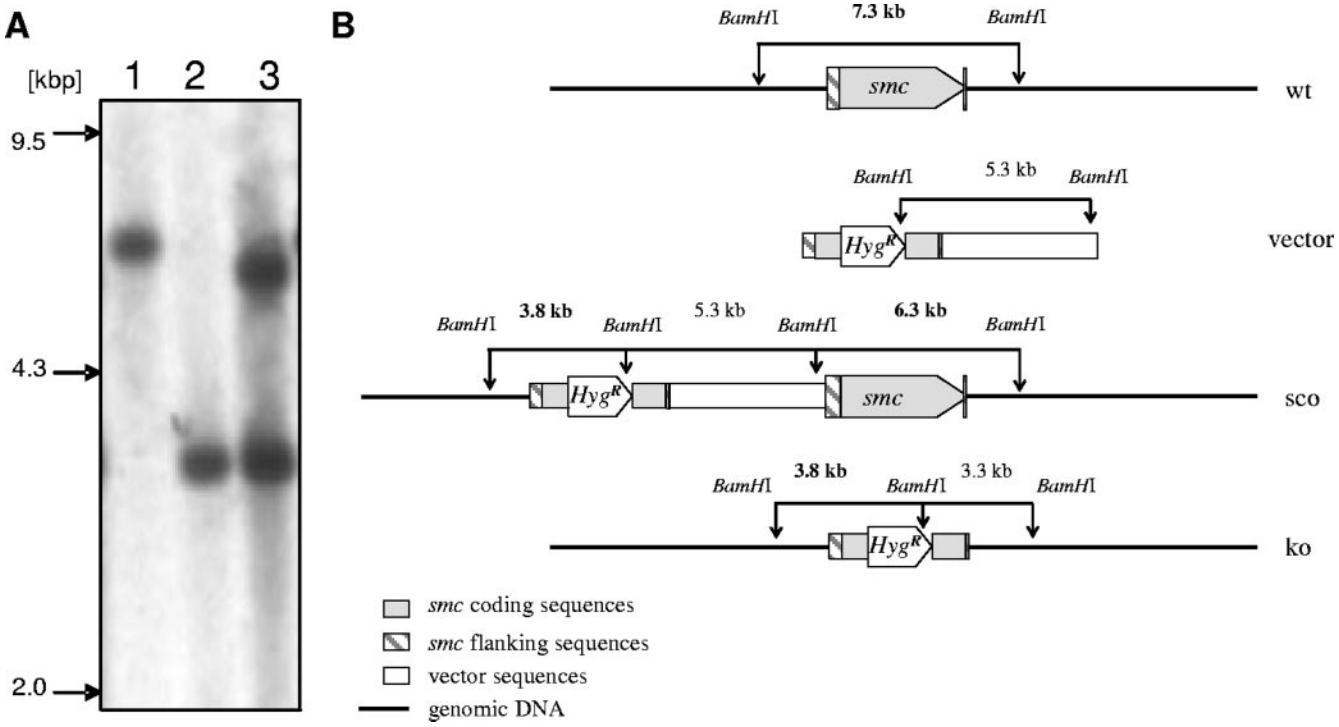


FIG. 1. Generation of *M. smegmatis smc* mutant. (A) Southern blot analysis. Genomic DNA from *M. smegmatis* wild type (lane 1), *smc* mutant (lane 2), and the *smc* single-crossover mutant (lane 3) was digested with BamHI and probed with a 685-bp PvuII gene fragment containing the 5′ flanking region of the *smc* gene. The presence of a single 3.8-kbp fragment in the mutant strain instead of a 7.3-kbp fragment observed in the parental strain demonstrates successful deletion of *smc* coding sequences. (B) Schematic illustration of the *smc* locus and the Southern blot strategy. Shown are the genomic organization of the wild type (wt), the knockout vector that contains the deletion allele *smc*::*hyg* (vector), the single-crossover genotype (sco), and the mutated genomic *smc* region in the knockout mutant (ko). Fragments detected by the probe specific for the 5′ flanking region are shown in bold letters.

other proteins, and SMC of *B. subtilis* was found to function in concert with two proteins called ScpA and ScpB (11). Both *M. tuberculosis* and *M. smegmatis* have homologues of these interacting proteins, suggesting that mycobacterial SMC is functional (ScpA homologues, Rv1709/MSMEG3748; ScpB homologues, Rv1710/MSMEG3749; sequences were obtained from The Institute for Genomic Research website and TubercuList of Institut Pasteur, respectively).

To study the function of mycobacterial SMC, we generated an *smc*-null mutant in *Mycobacterium smegmatis*, an established model for genetic and biochemical studies in mycobacteria (1, 14). The mutant was constructed by allelic-exchange mutagenesis (14, 23). Using genomic DNA as template, a 2-kb deletion allele was constructed by joining PCR products that contain parts of the ORF flanked by upstream and downstream sequences. The resulting deletion allele lacks bp 747 to 2660 of the 3,588-bp ORF. To allow proper selection of gene deletion mutants, a hygromycin resistance cassette was inserted between the two fragments, resulting in the deletion allele *smc*::*hyg*. Transformation and counterselection procedures for *M. smegmatis* were performed at room temperature. Transformants were plated on medium containing hygromycin to select for homologous recombination of the knockout vector into the genome. Southern blot analysis revealed a clone that displayed a pattern indicative of a single-crossover event, i.e., tandem arrangement of the deletion allele upstream of the functional *smc* allele (designated 5′ sco). The single-crossover clone was grown for 5 days in liquid medium at room temperature. Subsequently, the bacteria were plated on selective 7H10 agar containing streptomycin and hygromycin to select against maintenance of knockout vector sequences which contain the wild-type *rpsL* gene. Southern blot analyses demonstrated that in 11 of 24 analyzed clones, the wild-type *smc* gene was replaced by the deletion allele, resulting in *smc* knockout mutants (Fig. 1). All selection steps were set off with bacteria taken from frozen stocks to ensure that no suppressor mutations arose due to extensive passaging.

One prominent phenotype of previously described *smc* null mutants is temperature sensitivity (2, 15, 31). As many SMC mutants have distinctive growth defects at high growth rates (2), we analyzed the in vitro growth characteristics in 7H9 broth in comparison to the parental wild-type strain. 7H9 medium is known to support optimal in vitro growth of mycobacteria. In addition to growth at the standard incubation temperature (37°C), we assessed the growth rate at lower (30°C) and higher (42°C) temperatures. In contrast to *smc* null mutant phenotypes reported for *B. subtilis* and *C. crescentus*, the growth rates of the mutant and the wild type were similar in *M. smegmatis* at the various temperatures investigated (Fig. 2A). Doubling times in logarithmic phase were nearly identical, indicating that *smc* deletion does not affect in vitro growth of *M. smegmatis*. Additionally, the wild type and the mutant were indistinguishable with respect to colony morphology; when

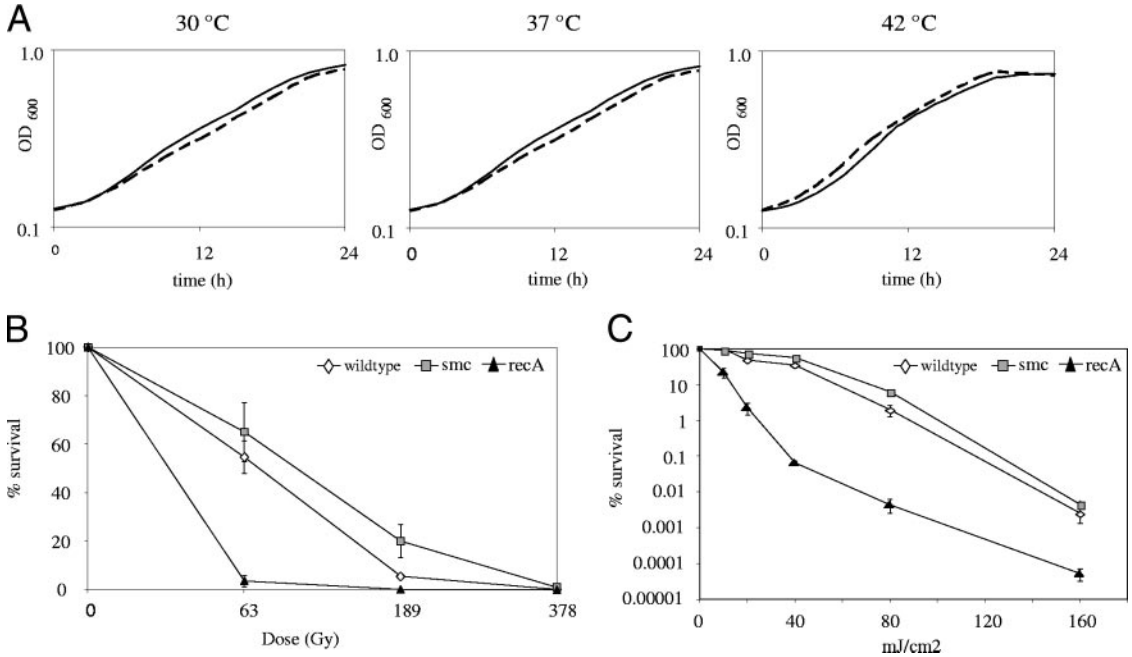


FIG. 2. Analysis of *M. smegmatis smc* mutant strain. (A) Growth of *M. smegmatis* wild-type strain (solid lines) and *smc* mutant strain (dashed lines) at various temperatures. OD_{600}, optical density at 600 nm. (B) Survival following exposure to ionizing radiation. (C) Survival following exposure to UV.

grown on 7H10 agar, both strains showed the typical features of mycobacterial colonies with rough dry surfaces and irregular edges (data not shown).

To assess if SMC deficiency leads to an increase in mutation frequency, the frequencies of spontaneous resistance to rifampin were determined. Four independent cultures of the parental wild-type strain and of the mutant strain were grown in 7H9 broth to mid-log phase, and aliquots were plated on 7H10 plates containing rifampin (175 μg/ml). To calculate the spontaneous mutation frequency, CFU were determined in parallel by plating serial dilutions on nonselective LB medium. Mutation frequencies were comparable in wild-type ($3.1 \times 10^{-8} \pm 1.5 \times 10^{-8}$) and mutant ($3.8 \times 10^{-8} \pm 1.0 \times 10^{-8}$; $P = 0.68$; $n = 4$; Student's t test) strains, indicating that SMC deficiency does not result in a mutator phenotype in *M. smegmatis*.

DNA gyrase, which generates negative supercoiling, and topoisomerase I, which prevents excessive negative supercoiling by DNA gyrase, are essential to maintain the optimal topological state of DNA in the cell (3). Previous investigations have indicated that SMC proteins mediate chromosome compaction by contributing to the introduction of negative supercoils into DNA (29). This is corroborated by the observation that *smc* and *mukB* mutants are hypersusceptible to gyrase inhibitors (25), while loss of MukB function can be partially compensated for by reducing topoisomerase I activity (29). To investigate if SMC interferes with DNA supercoiling in *M. smegmatis*, we determined MICs towards the gyrase inhibitor ofloxacin using the E-test (AB Biodisk). In contrast to findings made in other organisms (25), *M. smegmatis* wild-type and *smc* mutant strains displayed identical drug susceptibilities to ofloxacin, with MICs of 0.19 μg/ml.

SMC is known to participate in the repair of DNA double-strand breaks in various organisms (22, 27), prompting us to examine the ability of the SMC mutant strain to survive treatment with ionizing radiation. Mid-log-phase cultures were irradiated (0 to 378 Gy), and CFU were determined. As DNA double-strand breaks are also repaired by homologous recombination repair (13), we included a *recA* mutant strain as a control. In contrast to the *recA* mutant strain, no major difference in survival between the wild-type strain and the *smc* mutant strain was found (Fig. 2B). The eukaryotic SMC5/6 complex has been described to act in a second pathway for postreplicative repair of UV-induced lesions, which is different from the standard nucleotide excision repair pathway (18). In order to assess a possible role of mycobacterial SMC in repair of UV-induced lesions, we investigated the sensitivity of the SMC mutant strain to UV irradiation. Three independent cultures of each strain were grown in 7H9 medium until late log phase (optical density, 1 to 2). Serial dilutions of each strain were plated on LB agar. Subsequently, the plates were exposed to different UV doses (0 to 160 mJ/cm^2, 254 nm) and the number of CFU were determined. Compared to the *recA* mutant strain, which was impaired in survival, the wild-type strain and the *smc* mutant strain displayed similar susceptibilities to UV (Fig. 2C). Investigation of susceptibility to mitomycin C, a compound that damages DNA by generating intrastrand cross-links, revealed no difference between the *smc* mutant strain and the wild-type strain (data not shown). In summary, our data provide evidence that mycobacterial SMC has, if at all, a minor role in DNA repair.

Some genes are not essential for exponential growth but are prerequisites for remaining viable during stationary

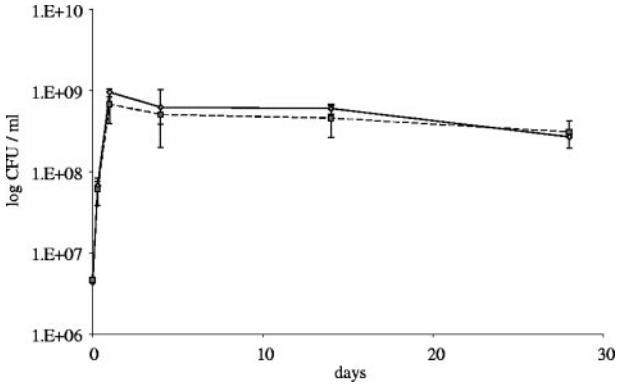


FIG. 3. Long-term survival of *M. smegmatis* wild-type strain (solid line) and *smc* mutant strain (dashed line).

phase (9). In order to gain insight into a possible role of SMC in long-term survival of *M. smegmatis*, we compared CFU counts of the *M. smegmatis* wild type and the *smc* mutant grown for 28 days in 7H9 broth. Viable counts of stationary cultures were similar for the two strains (Fig. 3), suggesting that SMC is not essential for long-term survival during stationary phase.

To study whether the findings obtained in the model organism *M. smegmatis* hold true for its pathogenic relatives, we generated an *M. tuberculosis smc* mutant. A deletion allele lacking 2,121 bp of the 3,618-bp *smc* ORF (the resulting mutant lacks bp 553 to 2674 of the ORF) was constructed by allelic-replacement mutagenesis as outlined for *M. smegmatis* above. Disruption of *smc* was verified by Southern blot analysis (Fig. 4). Growth of the mutant at 37°C in 7H9 broth supplemented with oleic acid-albumin-dextrose-catalase was not impaired, indicating that interruption of *smc* does not affect in vitro growth (data not shown). Survival in stationary phase was determined by comparing CFU counts of the *M. tuberculosis* wild-type and *smc* mutant strains grown for 83 days in 7H9 broth supplemented with oleic acid-albumin-dextrose-catalase. Both strains reach viable cell counts of about 1×10^8 CFU/ml after growth for 9 days, which for both strains dropped to approximately 3×10^7 CFU/ml after 83 days of incubation, suggesting that SMC is not required for long-term survival of *M. tuberculosis* (data not shown).

In this study *smc* mutant strains of *M. smegmatis* and *M. tuberculosis* were constructed. In contrast to previous data obtained from other bacterial *smc* null mutants, SMC deficiency did not result in apparent phenotypes. This is surprising given the crucial role of SMC proteins in chromosome dynamics and DNA repair in all organisms analyzed so far (12, 17, 19). Construction and analysis of *scpA* and *scpB* knockout mutants will be subjects of future investigations. These mutants eventually corroborate the results obtained with the *smc* mutants or reveal a possible additional, SMC-independent involvement of these proteins in mycobacterial chromosome organization.

Taken together, our observations suggest that maintenance of mycobacterial chromosome organization differs from that of previously described models. Further investigations of mycobacteria are likely to provide insights into new mechanisms which maintain genome integrity and ensure persistence.

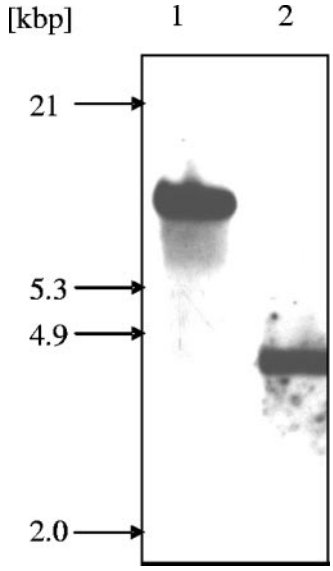


FIG. 4. Generation of *M. tuberculosis smc* mutant. Genomic DNA from *M. tuberculosis* wild-type (lane 1) and *smc* mutant (lane 2) strains was digested with BamHI and probed with a 935-bp NdeI gene fragment containing the 5′ flanking region of the *smc* gene. The presence of a single 4.3-kbp fragment instead of an 11.1-kbp fragment as seen in the parental strain demonstrates successful deletion of *smc* coding sequences.

This work was supported in part by grants from the University of Zurich (to B.S.), the Swiss National Science Foundation (to P.S. and E.C.B.; BO-3200-68488), and the European Community (to B.S. and E.C.B.; CSI_LTB, LSHP-CT-2007-037235).

REFERENCES

1. **Acharya, N., and U. Varshney.** 2002. Biochemical properties of single-stranded DNA-binding protein from *Mycobacterium smegmatis*, a fast-growing mycobacterium and its physical and functional interaction with uracil DNA glycosylases. J. Mol. Biol. **318:**1251–1264.
2. **Britton, R. A., D. C.-H. Lin, and A. D. Grossman.** 1998. Characterization of a prokaryotic SMC protein involved in chromosome partitioning. Genes Dev. **12:**1254–1259.
3. **Champoux, J. J.** 2001. DNA topoisomerases: structure, function and mechanism. Annu. Rev. Biochem. **70:**369–413.
4. **Cobbe, N., and M. M. Heck.** 2004. The evolution of SMC proteins: phylogenetic analysis and structural implications. Mol. Biol. Evol. **21:**332–347.
5. **Cobbe, N., and M. M. S. Heck.** 2000. Review: SMCs in the world of chromosome biology—from prokaryotes to higher eukaryotes. J. Struct. Biol. **129:**123–143.
6. **Cole, S. T., K. Eiglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M.-A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell.** 2001. Massive gene decay in the leprosy bacillus. Nature **409:**1007–1011.
7. **Dervyn, E., M. F. Noirot-Gros, P. Mervelet, S. McGovern, S. D. Ehrlich, P. Polard, and P. Noirot.** 2004. The bacterial condensin/cohesin-like protein complex acts in DNA repair and regulation of gene expression. Mol. Microbiol. **51:**1629–1640.
8. **Frenkiel-Krispin, D., and A. Minsky.** 2006. Nucleoid organization and the maintenance of DNA integrity in *E. coli*, *B. subtilis* and *D. radiodurans*. J. Struct. Biol. **156:**311–319.
9. **Hampshire, T., S. Soneji, J. Bacon, B. W. James, J. Hinds, K. Laing, R. A. Stabler, P. D. Marsh, and P. D. Butcher.** 2004. Stationary phase gene expression of *Mycobacterium tuberculosis* following a progressive nutrient depletion: a model for persistent organisms? Tuberculosis **84:**228–238.

10. **Hirano, M., and T. Hirano.** 2002. Hinge-mediated dimerization of SMC protein is essential for its dynamic interaction with DNA. EMBO J. **21:**5733–5744.
11. **Hirano, M., and T. Hirano.** 2004. Positive and negative regulation of SMC-DNA interactions by ATP and accessory proteins. EMBO J. **23:**2664–2673.
12. **Hirano, T.** 2005. SMC proteins and chromosome mechanics: from bacteria to humans. Philos. Trans. R. Soc. Lond. B Biol. Sci. **360:**507–514.
13. **Ishino, Y., T. Nishino, and K. Morikawa.** 2006. Mechanisms of maintaining genetic stability by homologous recombination. Chem. Rev. **106:**324–339.
14. **Jacobs, W. R., Jr., G. V. Kalpana, J. D. Cirillo, L. Pascopella, S. B. Snapper, R. A. Udani, W. Jones, R. G. Barletta, and B. R. Bloom.** 1991. Genetic systems for mycobacteria. Methods Enzymol. **204:**537–555.
15. **Jensen, R. B., and L. Shapiro.** 1999. The *Caulobacter crescentus smc* gene is required for cell cycle progression and chromosome segregation. Proc. Natl. Acad. Sci. USA **96:**10661–10666.
16. **Jessberger, R.** 2002. The many functions of SMC proteins in chromosome dynamics. Nat. Rev. Mol. Cell Biol. **3:**767–778.
17. **Lehmann, A. R.** 2005. The role of SMC proteins in the responses to DNA damage. DNA Repair **4:**309–314.
18. **Lehmann, A. R., M. Walicka, D. J. Griffiths, J. M. Murray, F. Z. Watts, S. McCready, and A. M. Carr.** 1995. The *rad18* gene of *Schizosaccharomyces pombe* defines a new subgroup of the SMC superfamily involved in DNA repair. Mol. Cell. Biol. **15:**7067–7080.
19. **Losada, A., and T. Hirano.** 2005. Dynamic molecular linkers of the genome: the first decade of SMC proteins. Genes Dev. **19:**1269–1287.
20. **Melby, T. E., C. N. Ciampaglio, G. Briscoe, and H. P. Erickson.** 1998. The symmetrical structure of structural maintenance of chromosomes (SMC) and MukB proteins: long, antiparallel coiled coils, folded at a flexible hinge. J. Cell Biol. **142:**1595–1604.
21. **Nasim, A., and B. P. Smith.** 1975. Genetic control of radiation sensitivity in *Schizosaccharomyces pombe*. Genetics **79:**573–582.
22. **Potts, P. R., M. H. Porteus, and H. Yu.** 2006. Human SMC5/6 complex promotes sister chromatid homologous recombination by recruiting the SMC1/3 cohesin complex to double-strand breaks. EMBO J. **25:**3377–3388.
23. **Sander, P., A. Meier, and E. C. Böttger.** 1995. *rpsL*+: a dominant selectable marker for gene replacement in mycobacteria. Mol. Microbiol. **16:**991–1000.
24. **Sandman, K., S. L. Pereira, and J. N. Reeve.** 1998. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. Cell. Mol. Life Sci. **54:**1350–1364.
25. **Sawitzke, J. A., and S. Austin.** 2000. Suppression of chromosome segregation defects of *Escherichia coli muk* mutants by mutations in topoisomerase I. Proc. Natl. Acad. Sci. USA **97:**1671–1676.
26. **Schär, P., M. Fasi, and R. Jessberger.** 2004. SMC1 coordinates DNA double-strand break repair pathways. Nucleic Acids Res. **32:**3921–3929.
27. **Sjogren, C., and K. Nasmyth.** 2001. Sister chromatid cohesion is required for postreplicative double-strand break repair in *Saccharomyces cerevisiae*. Curr. Biol. **11:**991–995.
28. **Soppa, J.** 2001. Prokaryotic structural maintenance of chromosomes (SMC) proteins: distribution, phylogeny, and comparison with MukBs and additional prokaryotic and eukaryotic coiled-coil proteins. Gene **278:**253–264.
29. **Tadesse, S., and P. L. Graumann.** 2006. Differential and dynamic localization of topoisomerases in *Bacillus subtilis*. J. Bacteriol. **188:**3002–3011.
30. **Tufariello, J. M., J. Chan, and J. L. Flynn.** 2003. Latent tuberculosis: mechanisms of host and bacillus that contribute to persistent infection. Lancet Infect. Dis. **3:**578–590.
31. **Volkov, A., J. Mascarenhas, C. Andrei-Selmer, H. D. Ulrich, and P. L. Graumann.** 2003. A prokaryotic condensin/cohesin-like complex can actively compact chromosomes from a single position on the nucleoid and binds to DNA as a ring-like structure. Mol. Cell. Biol. **23:**5638–5650.

Stabilization of the genome of the mismatch repair deficient *Mycobacterium tuberculosis* by context-dependent codon choice

Roger M. Wanner[1], Carolin Güthlein[1], Erik C. Böttger[1], Burkhard Springer[2], Martin Ackermann[3]

[1] Institut für Medizinische Mikrobiologie, Universität Zürich, Gloriastrasse 30/32, CH-8006 Zurich, Switzerland

[2] Institute of Medical Microbiology and Hygiene Graz, Austrian Agency for Health and Food Safety, Beethovenstrasse 6, A-8010 Graz, Austria

[3] Institute of Integrative Biology, ETH Zurich, Universitaetsstrasse 16, CH-8092 Zurich, Switzerland

Correspondence: Martin.Ackermann@env.ethz.ch

Abstract

Background: Microbial pathogens frequently evade host defense mechanisms by stochastically varying phenotypic traits. One major source of this variability are hypermutable simple sequence repeats in the genomes of these pathogens. However, such unstable sequences also come at a cost, as mutations are often deleterious. Here, we analyzed how these opposing forces shaped genome stability in the human pathogen *Mycobacterium tuberculosis*. *M. tuberculosis* lacks a mismatch repair system, and this renders simple sequence repeats particularly unstable.

Results: We found that protein coding genes of *M. tuberculosis* do not contain long mono-, di-nucleotide repeats that could be interpreted as mutational hotspots. Rather, proteins are encoded by using codons in a way that prevents the emergence of these unstable sequences. This context-dependent codon choice leads to a strong decrease in the estimated frame-shift mutation rate and thus to an increase in genome stability.

Conclusions: These results indicate that a context-specific codon choice helps to maintain genome integrity in this mismatch repair deficient pathogen.

Background

M .tuberculosis is the causative agent of tuberculosis. It is one of the most harmful pathogens causing 10 million new infections and 1.6 million deaths every year [1]. There is a considerable interest in understanding the genome evolution of this pathogen, as this can further our understanding of pathogenicity and contribute to the development of new vaccines [2-5]. Here we focus on one aspect of genome evolution that plays a pivotal role in the infection biology of many pathogens: structural properties of the DNA sequence that determine the local mutation rate.

The rate at which a stretch of DNA mutates is influenced by the nucleotide sequence itself. Certain sequences are inherently prone to errors during replication and gene expression, while other sequences are more stable. Particularly unstable are simple sequence repeats, which consist of short motifs of up to six nucleotides that are repeated several times [6]. Simple sequence repeats are prone to length variation during replication because of DNA-polymerase slippage [7]. The mutation rate in simple sequence repeats is much higher than in non-repeated sequences [8, 9], and increases with the number of repeats [7, 10-13]. If simple sequence repeats are located in coding regions, and if they consist of motifs whose length is not a multiple of three, then a change in the number of repeats leads to a frame-shift mutation, and thereby to a complete loss of the amino acid sequence information.

Some organisms, and particularly microbial pathogens, contain conspicuously long simple sequence repeats in coding regions [14]. For example, the bacteria *Haemophilus influenzae* [15], *Neisseria meningitidis* [16]*, and Campylobacter jejuni* [17] contain mononucleotide, dinucleotide and tetranucleotide repeats in genes that are involved in interactions with the host. The most plausible biological function of these repeats is that they promote phenotypic variation among otherwise clonal cells by altering protein expression profiles at a high rate [18]. This phenotypic variation might allow pathogens to evade the immune system, as well as to increase the probability that a fraction of a clonal population survives changing conditions [19].

However, unstable nucleotide sequences also come at a cost, as many mutational changes are detrimental. Recent studies suggested that simple sequence repeats are avoided in genomes of non-pathogenic organisms [20-22]. It is thus currently not clear whether unstable nucleotide sequences are ubiquitous among microbial pathogens, or whether selection for stability usually also prevails in these genomes and leads to a bias against simple sequence repeats. Given the significance of *M. tuberculosis* as pathogen, it is important to investigate whether the genome of this microorganism is biased towards stability or instability.

Mycobacteria are especially prone to polymerase slippage in simple sequence repeats, for two reasons. First, their high G+C content (65.6% in *M. tuberculosis*; [2]) makes the emergence of repeats more likely. Second, they lack a mismatch repair system (MMR) [23] or other orthologous repair systems that could compensate for this deficiency [2, 24, 25]. MMR deficiency leads to a high mutation rate in simple sequence repeats, and especially in mononucleotide repeats [26], and it has been shown experimentally that in mycobacteria, the frame-shift rate in mononucleotide repeats is particularly high [23].

Here, we analyzed whether proteins in *M. tuberculosis* are encoded in a way that prevents or promotes the emergence of unstable sequences. A recent study has analyzed the distribution of simple sequence repeats in mycobacterial genomes, and reported that long repeats were less frequent than expected if nucleotides were randomly distributed [27]. We investigated here the causes and consequences of this under-representation. Focusing on short nucleotide repeats, which are a particularly important determinant of stability, we found that a context-dependent codon choice in the genome of *M. tuberculosis* leads to a bias against these unstable sequences, and that this leads to a strong increase in estimates of genetic stability.

Results and Discussion

Comparing frequencies of mononucleotide repeats in coding regions and intergenic regions.

A first insight into the distribution of short nucleotide repeats can be gained from the comparison of coding and intergenic regions. The frequency of repeats in a particular region of the genome is determined by the combination of mutational processes and selection. Mutational processes are expected to be similar in coding and intergenic regions (except for transcription-mediated mutations and repair [28]). In contrast, selection is generally stronger on coding regions than on intergenic regions. If selection leads to a bias for or against repeats, one would expect this bias to be stronger in coding regions than in intergenic regions. We tested this by comparing observed and expected occurrences of repeats in these two regions. We focused on mononucleotide repeats, because they are particularly unstable and more frequent than other types of repeats.

We first determined the number of mononucleotide repeats that would be expected if nucleotides were randomly distributed within each region, and then compared this number to the actually observed frequencies of repeats. We found that repeats of three nucleotides and more are less frequent than expected in both coding and intergenic regions (Figure 1). The bias increases with increasing repeat length, and is stronger in coding than in intergenic regions (at $p<0.0001$ for repeats of length 3 to 7;

chi-square test). The latter finding suggests that the bias against repeats might be a consequence of selection against unstable sequences.

Context-dependent codon choice in *M. tuberculosis* limits the emergence nucleotide repeats.

In this first comparison, we calculated the expected frequencies of mononucleotide repeats under the assumption that nucleotides are randomly distributed. This null-model is appropriate for comparisons involving intergenic regions. In coding regions, however, a null-model that preserves the amino acid sequence and gene-specific codon frequencies is more adequate [22]. Such a modified null-model accounts for constraints on the DNA sequence imposed by the need to encode a particular protein, and it accounts for regional variation in GC-content and codon usage. We thus re-analyzed coding regions by using this modified null-model. For each gene, we determined the codon frequencies, and then re-arranged the codons while preserving the amino acid sequence. Such a re-arrangement is possible because the genetic code is degenerate; as most amino acids are encoded for by more than one codon, a given amino acid sequence can be encoded for by different nucleotide sequences. For each gene, we generated 100 randomized sequences. Then we asked whether the observed nucleotide sequence differs from the random sequences in the frequency of short nucleotide repeats, and thus in terms of stability.

We again first focused on mononucleotide repeats, and found that these repeats are strongly under-represented in coding regions of *M. tuberculosis* (Figure 2). The under-representation increases with increasing repeat length. Furthermore, the bias is stronger against runs of cytosine and guanine than against runs of adenine and thymine. This pattern can be explained by the fact that mononucleotide repeats of cytosine and guanine have a higher mutation rate than repeats of adenine and thymine [22, 29, 30].

Importantly, the bias against mononucleotide repeats depicted in Figure 2 is not a consequence of an avoidance of certain codons; the randomized sequences have the same codon frequencies as the observed sequence. Rather, the bias results from a context-dependent codon choice. For example, the amino acid proline can be encoded by CCA, CCC, CCG and CCT. The codon CCC is used often (at 34%) at positions that are *not* followed by a cytosine. At positions followed by one cytosine, the frequency of CCC drops to 16%. With every additional cytosine following, the frequency decreased further (Figure 3). This decrease is statistically significant ($p<0.0001$, logistic regression). This same pattern holds for the codons AAA and GGG (and is also significant at $p<0.0001$; logistic regression). For TTT, the pattern is slightly different; the relative frequency of TTT first increases at positions followed by one thymidine, and then decreases at positions followed by more than one thymidine.

Quantifying the stabilizing effect of context-dependent codon choice

As nucleotide repeats are prone to frame-shift mutations, the under-representation of mononucleotide repeats in the genome of *M. tuberculosis* is expected to lead to a decrease in the mutation rate. We sought to quantify the magnitude of this effect. First, we estimated the frame-shift mutation rate in mononucleotide repeats in coding regions of the *M. tuberculosis* genome. This estimate was based on published measurements of the frame-shift mutation rate in mononucleotide repeats in *Mycobacteria* and other bacteria (see Methods). Then, we estimated the frame-shift mutation rate that would be expected if codons were used at random, irrespective of the local context. To do so, we analyzed the randomized genomes, where codons were used in a context-independent manner. The ratio between estimated frame-shift mutation rates in the real and in the randomized genomes gave us an estimate for the magnitude of the stabilizing effect of context-dependent codon choice in the genome of *M. tuberculosis*.

This ratio was sensitive to the mutation rate parameters retrieved from the literature, and we thus carried out the analysis for a range of parameter values. This analysis showed that the stabilizing effect of context-dependent codon choice was surprisingly strong: for a plausible range of biological parameters, the estimated frame-shift mutation rate in mononucleotide repeats in the randomized sequences was between 10 and 100 times higher than in the real sequence (Figure 4). For the parameter values that we considered most plausible (see Methods), the randomized genomes had an estimated frame-shift mutation rate that was about 20 times higher than the estimate for the real genome.

In fact, without context dependent codon choice, one would expect that frame-shift mutations in mononucleotide repeats would be the most frequent type of mutations in the *M. tuberculosis* genome. For the most plausible parameter values, we estimated that in the absence of context-dependent codon choice, the probability to acquire a frame-shift mutation in a mononucleotide repeat in a protein-coding gene would be 3% per replication (see Additional File 2). This value is about ten times higher than the *total* number of mutations expected per genome and replication in DNA-based microbes [31]. This means without bias against mononucleotide repeats, frame-shift mutations would happen very frequently, and they would constitute a very large fraction of all mutations that occur. This indicates that context-dependent codon choice has a substantial impact on genome stability of *M. tuberculosis*.

We carried out the same analysis for the genome of the non-pathogenic *Escherichia coli* strain K12. Previous studies reported a bias against mononucleotide repeats for this organism, and attributed this bias to selection for stability (see [22]). Interestingly, the estimated stabilizing effect of context-

dependent codon choice was much smaller for *E. coli* than for *M. tuberculosis* (Fig. 4). For the combination of parameters that we considered most plausible for *M. tuberculosis*, the estimated mutation rates for the real and the randomized genomes of *E. coli* differed by less than a factor of three (compared to a factor of 20 for *M. tuberculosis*). One possible explanation for this difference is that in the mismatch-deficient *M. tuberculosis*, selection against unstable sequences is stronger than in *E. coli*, and that this results in a stronger stabilization of the genome by context-dependent codon choice.

Observed and expected frequencies of dinucleotide and tetranucleotide repeats

In addition to mononucleotide repeats, di- and tetranucleotide repeats are other important determinants of the local mutation rate; they are also involved in phase variation in a number of pathogenic bacteria [32]. It is thus important to investigate whether context-dependent codon choice also leads to a bias against these two other types of repeats. To determine the expected frequencies of di- and tetranucleotide repeats, we again used a null model that conserved the amino acid sequence and the gene-specific codon frequencies. Then, we compared these expected frequencies of di- and tetranucleotide repeats to the actually observed frequencies.

Dinucleotide repeats are generally rare in the genome of *M. tuberculosis,* and the observed numbers are close to expected. There are twelve different types of dinucleotides (discounting AA, CC, GG and TT, which are also mononucleotide repeats). The longest observed dinucleotide repeats are of length five; there is total of 41 occurrences of AC_5, CA_5, CG_5 and GC_5. By analyzing all dinucleotide repeats of length three and longer, we detected 15 cases where the observed number is significantly different from the expectation; in 11 of these 15 cases, the expected number exceeds the observed number. In the remaining four cases, the expected number is lower than the observed number (see Additional File 1).

The results for tetranucleotide repeats are similar. There are 240 tetranucleotides that are not also mono- or dinucleotide repeats. The longest observed tetranucleotide repeats are of length four (two occurrences of $GGCC_4$). There are only 88 cases where the observed frequency of tetranucleotide repeats (of length two and longer) is significantly different from the expectation, and in 51 of those the expected number exceeds the observed number (see Additional File 1). These results indicate that context-dependent codon choice leads to some bias against di- and tetranucleotide repeats, but this bias is small compared to mononucleotide repeats. This might be because mononucleotide repeats tend to have higher frame-shift mutation rates than other repeats, and tend to be more de-stabilized by the absence of a mismatch repair system [33].

Selection or a neutral process?

The under-representation of short nucleotide repeats in the genome of *M. tuberculosis* could either result from selection for stable nucleotide sequences, or it could result from a neutral process [34]. A candidate for a neutral process is a deletion bias in nucleotide repeats. If deletions in short nucleotide repeats are more frequent than insertions, then repeats would decrease in size over the course of generations, and long repeats would become rare. However, in mono-, di- and tetranucleotide repeats, most deletions result in frame-shift mutation, which are usually deleterious; only deletions of a multiple of three units retain the reading frame. As most mutations in nucleotide repeats involve only one or two units [33], a deletion bias might not be a likely explanation for the under-representation of repeats observed here.

Another observation is consistent with the idea that the observed pattern is a consequence of selection: the bias against repeats is more pronounced in coding than in non-coding regions, as would be expected if it resulted from selection against targets of frame-shift mutations. There was also evidence for the action of selection in a previous study with other organisms that investigated statistical associations between the occurrence of repeats and the expression level and essentiality of genes [22].

However, the two types of explanations are not mutually exclusive, and they might both be involved in the observed pattern. And both explanations have the same final consequences: they lead to a genome that has fewer short nucleotide repeats and that is thus more stable. Interestingly, one would expect both types of processes to operate more strongly in organisms that lack an efficient mismatch repair system; in such organisms, selection for stable sequences would be stronger, and deletion biases would operate more quickly, thereby eliminating targets for frame-shift mutations. As a consequence, the lack of a mismatch repair system might, over evolutionary times, lead to genomes that partially compensate for this lack by an increased structural stability.

Conclusions

In the present study we introduced the concept that a context-dependent codon choice can be an alternative to a mismatch repair system for attaining genetic stability. We showed that in the pathogenic, MMR-deficient *M. tuberculosis,* codons are used in a context-specific way that limits the emergence of short nucleotide repeats, and therefore reduces the number of targets for frameshift mutations. Context-dependent codon choice leads to a strong decrease in the estimated frame-shift mutation rate. This structural stability in *M. tuberculosis* is more pronounced than in the mismatch repair-proficient bacterium *E. coli*. This is consistent the idea that a deficiency in enzymatic DNA repair increases the selection for inherently stable genomes.

It is interesting to note that *M. tuberculosis* is an obligate pathogen, while *E. coli* is free-living. Genomes of bacterial pathogens often contain long mononucleotide repeats in genes involved in

interactions with the host [14]. However, the strong bias against short nucleotide repeats in *M. tuberculosis* is in line with the idea that in all organisms there is a core set of genes that are under selection for fulfilling their function in a stable and ordered manner [32], and suggest that in *M. tuberculosis* a dominant fraction of genes belongs to this group.

Methods

Origin of data

Sequence data of the whole genome and intergenic regions of *M. tuberculosis* H37Rv (genebank number AL123456) were retrieved from http://tigr.org. Sequence data of *E. coli* K12 were retrieved from NCBI (accession no. 000913).

Comparing mononucleotide repeats in coding and intergenic regions with a null-model that assumes random distribution of nucleotides

To compare the distribution of mononucleotide repeats in coding and intergenic regions, we used a null-model based on a random distribution of nucleotides within each compartment. We first determined the observed distribution of mononucleotide repeats for each nucleotide, as well as the frequency of each nucleotide in both compartments. Then we calculated the expected distribution of mononucleotide repeats in both compartments with a permutation model. For example: the expected number of G_6 repeats is determined by the probability of finding exactly 6 consecutive Gs by chance. As long as a compartment is reasonably large, this is very closely approximated by $((N-G)^2/N^2) \times (G^6/N^6)N$, where G is the number of Gs and N is the total number of all nucleotides within this compartment.

For each combination of nucleotide and repeat length, we then tested whether the proportions of expected and observed numbers differed between coding and intergenic regions. Observed and expected numbers in coding and intergenic regions were regarded as entries in a 2*2 contingency table, and a chi-square test was used to test for an association between the values for observed/expected and coding/noncoding.

Comparing observed and expected numbers of mononucleotide repeats in coding regions with a null-model that preserves the amino acid sequence and the codon frequencies

The expected distribution of mononucleotide repeats in coding regions was calculated with a null-model that preserved the amino acid sequence and the ORF-specific codon frequencies. For each ORF, we first determined the codon frequencies and the amino acid sequence. Then, we generated 1000 random nucleotide sequences that preserved these two features. In these randomized sequences, the number and length of all mononucleotide repeats was determined. This data was used to calculate the

average expected number of repeats, as well as the distribution around this average, for every combination of nucleotide and length. We also determined the percentile range of 2.5–97.5 of this distribution. If the observed number of mononucleotide repeats fell outside this range, we concluded that this observation was not consistent with the null-model; this indicated that repeats of this nucleotide and this particular length were significantly over- or underrepresented. For Fig. 2, the observed number of repeats as well as the 2.5–97.5 percentile range was standardized by dividing by the average of the expected number (for details see [22]).

Determining the context-dependent usage of homogenous codons

In all ORFs we determined the occurrences of the four amino acids that can be encoded by a homogeneous codon (i.e., a codon that consists of three identical nucleotides) as well as one or more non-homogeneous codons (codons that do not consist of three identical nucleotides). These are phenylalanine (TTT and TTC), proline (CCA, CCC, CCG, and CCT), lysine (AAA and AAC), and glycine (GGA, GGC, GGG, and GGT). For each occurrence of one of these amino acids, we determined whether the codon was homogeneous or non-homogeneous, and whether it was followed by one or more of the nucleotides that constitute the homogeneous codon. This data was used to determine how the usage of homogeneous codons changed at positions immediately followed by one or more nucleotides of the same type. We used logistic regression to test whether the probability for a homogeneous codon changed with the number of nucleotides of the same type following. The program Jmp (version 6, SAS Institute Inc.) was used for logistic regression.

Estimating the frameshift mutation rate

To estimate the frameshift mutation rates in mononucleotide repeats composed of different nucleotides and of different length, we used estimates of biological parameters from the literature. The first assumption was that the frame-shift mutation rate in a C_6 repeat in Mycobacteria is about $3*10^7$ [23]. The second assumption was that in bacteria, the frame-shift mutation rate in repeats of C and G is about five times higher than in repeats of A and T of the same length [35]. The third assumption was that in MMR deficient bacteria, the mutation rate in mononucleotide repeats increases by about a factor three with each additional nucleotide that is added to the repeat [36]. A further assumption was that the frame-shift mutation rate in mononucleotide 'repeats' of length two and three (for example, CC and CCC) is negligible, and that only repeats of length four and longer have a significant frame-shift mutation rate.

The estimate of the genome-wide frame-shift mutation rate was particularly sensitive to the last two assumptions. For this reason, we varied these assumptions to obtain estimates of the frame-shift mutation rate for a whole range of parameter values (Figure 4). This lead to the following equations for estimated frame-shift mutation rates $U_{FS(n)}$ in a mononucleotide repeat of length $n \geq i$ (where i is

the minimal length that has a significant frame-shift mutation rate; mononucleotide repeats with a length smaller than *i* were assumed to have a frame-shift mutation rate of zero):

$U_{FS(n)} = 3*10^{-7}/b*c^{n-6}$

where *b*, a parameter that accounts for differences in the mutation rate between repeats of A and T versus repeats of C and G, is equal to five for the nucleotides A and T and equal to one for C and G, and *c* is the factor by which the mutation rate increases with each additional nucleotide. The factor $3*10^{-7}$ is based on the estimated frame-shift mutation rate in C_6 repeats of Mycobacteria [23]. This equation was used to calculate the estimated frame-shift mutation rate for every mononucleotide repeat in the real genome. We then added those rates to obtain the estimate of the total frame-shift mutation rate per genome. We used the same method to determine the estimated frame-shift mutation rate in the randomized genomes. Based on published data (see above), we concluded that the parameter values *c*=3 and *i*=4 are most plausible. For Figure 4, the parameter *c* was varied from 2 to 4, and the parameter *i* was varied from 4 to 6.

Comparing observed and expected numbers of di- and tetranucleotide repeats in coding regions with a null-model that preserves the amino acid sequence and the codon frequencies

To determine expected frequencies of di- and tetranucleotide repeats, we used the same null-model as for mononucleotide repeats. We generated 100 random nucleotide sequences of the coding regions that preserved the amino acid sequence of each ORF and the ORF-specific codon frequencies. In these randomized sequences, the number and length of all di-and tetranucleotide repeats was determined. This data was used to calculate the average expected number of repeats, as well as the percentile range of 2.5–97.5, for every combination of nucleotide and length.

Authors's contributions

All authors conceived of the study. MA and RMW performed the analysis. MA and RMW wrote the manuscript. ECB and BS edited and improved the manuscript.

Acknowledgments

We thank Ralph Schlapbach and Ulrich Wagner from the Functional Genomics Center Zurich for inspiring discussions at the beginning of this project. This work was supported by research grants from the University of Zurich to B.S. and from the Swiss National Science Foundation to E.C.B. M.A. was supported by the Swiss National Science Foundation, by the Roche Research Foundation, and by the Novartis Foundation.

References

1. WHO: **Tuberculosis Fact Sheet No. 104**. In. Geneva: World Health Organization; 2007.
2. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry Iii CE: **Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence**. *Nature* 1998, **393**:537-544.
3. Jordan IK, Makarova KS, Spouge JL, Wolf YI, Koonin EV: **Lineage-Specific Gene Expansions in Bacterial and Archaeal Genomes**. *Genome Research*.
4. Ho TBL, Robertson BD, Taylor GM, Shaw RJ, Young DB: **Comparison of Mycobacterium tuberculosis genomes reveals frequent deletions in a 20 kb variable region in clinical isolates**. *Yeast* 2000, **17**(4):272-282.
5. Brosch R, Gordon SV, Buchrieser C, Pym AS, Garnier T, Cole ST: **Comparative genomics uncovers large tandem chromosomal duplications in Mycobacterium bovis BCG Pasteur**. *Yeast* 2000, **17**(2):111-123.
6. Schlötterer C: **Evolutionary dynamics of microsatellite DNA**. *Chromosoma* 2000, **109**(6):365-371.
7. Tran HT, Keen JD, Kricker M, Resnick MA, Gordenin DA: **Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants**. *Molecular and Cellular Biology* 1997, **17**(5):2859-2865.
8. Jin L, Macaubas C, Hallmayer J, Kimura A, Mignot E: **Mutation rate varies among alleles at a microsatellite locus: Phylogenetic evidence**. *Proc Natl Acad Sci US A* 1996, **93**(26):15285-15288.
9. Schlotterer C: **High mutation rate of a long microsatellite allele in Drosophila melanogaster provides evidence for allele-specific mutation rates**. In., vol. 15: SMBE; 1998: 1269-1274.
10. Yamada NA, Smith GA, Castro A, Roques CN, Boyer JC, Farber RA: **Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells**. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 2002, **499**(2):213-225.
11. Strand M, Prolla TA, Liskay RM, Petes TD: **Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair**. *Nature* 1993, **365**(6443):274-276.
12. Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD: **Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes**. *Mol Cell Biol* 1997, **17**(285):1-2858.
13. Hanford MG, Rushton BC, Gowen LC, Farber RA: **Microsatellite mutation rates in cancer cell lines deficient or proficient in mismatch repair**. *Oncogene* 1998, **16**(18):2389-2393.
14. Moxon ER, Rainey PB, Nowak MA, Lenski RE: **Adaptive Evolution of Highly Mutable Loci in Pathogenic Bacteria**. *Current Biology* 1994, **4**(1):24-33.
15. Hood DW, Deadman ME, Jennings MP, Bisercic M, Fleischmann RC, Venter JC, Moxon ER: **DNA repeats identify novel virulence genes in Haemophilus influenzae**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(20):11121-11125.
16. Saunders NJ, Jeffries AC, Peden PF, Hood DW, Tettelin H, Rappuoli R, Moxon ER: **Repeat-associated phase variable genes in the complete genome sequence of Neisseria meningitidis strain MC58**. *Molecular Microbiology* 2000, **37**(1):207-215.
17. Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S *et al*: **The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences**. *Nature* 2000, **403**(6770):665-668.
18. Hallet B: **Playing Dr Jekyll and Mr Hyde: combined mechanisms of phase variation in bacteria**. *Current Opinion in Microbiology* 2001, **4**(5):570-581.
19. van der Woude MW: **Re-examining the role and random nature of phase variation**. *FEMS Microbiol Lett* 2006, **254**:190-197.
20. Metzgar D, Bytof J, Wills C: **Selection against frameshift mutations limits microsatellite expansion in coding DNA**. *Genome Research* 2000, **10**(1):72-80.

21. Borstnik B, Pumpernik D: **Tandem Repeats in Protein Coding Regions of Primate Genes**. *Genome Res* 2002, **12**(6):909-915.
22. Ackermann M, Chao L: **DNA sequences shaped by selection for stability**. *Plos Genetics* 2006, **2**(2):224-230.
23. Springer B, Sander P, Sedlacek L, Hardt WD, Mizrahi V, Schar P, Bottger EC: **Lack of mismatch correction facilitates genome evolution in mycobacteria**. *Molecular Microbiology* 2004, **53**(6):1601-1609.
24. Mizrahi V, Andersen SJ: **DNA repair in Mycobacterium tuberculosis. What have we learnt from the genome sequence**. *Molecular Microbiology* 1998, **29**(6):1331-1339.
25. Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR, Wheeler PR, Honoré N, Garnier T, Churcher C, Harris D: **Massive gene decay in the leprosy bacillus**. *Nature* 2001, **409**:1007-1011.
26. Modrich P, Lahue R: **Mismatch Repair in Replication Fidelity, Genetic Recombination, and Cancer Biology**. *Annual Review of Biochemistry* 1996, **65**(1):101-133.
27. Sreenu VB, Kumar P, Nagaraju J, Nagarajaram HA: **Simple sequence repeats in mycobacterial genomes**. *Journal of Biosciences* 2007, **32**(1):3-15.
28. Francino MP, Chao L, Riley MA, Ochman H: **Asymmetries generated by transcription-coupled repair in enterobacterial genes**. *Science* 1996, **272**(5258):107-109.
29. Denver DR, Morris K, Kewalramani A, Harris KE, Chow A, Estes S, Lynch M, Thomas WK: **Abundance, distribution, and mutation rates of homopolymeric nucleotide runs in the genome of Caenorhabditis elegans**. *Journal of Molecular Evolution* 2004, **58**(5):584-595.
30. Harfe BD, Jinks-Robertson S: **Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in Saccharomyces cerevisiae**. *Genetics* 2000, **156**(2):571-578.
31. Drake JW: **A constant rate of spontaneous mutation in DNA-based microbes**. *Proceedings of the National Academy of Sciences of the United States of America* 1991, **88**(16):7160.
32. Moxon R, Bayliss C, Hood D: **Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation**. *Annu Rev Genet* 2006, **40**:307–333.
33. Jacob KD, Eckert KA: **Escherichia coli DNA polymerase IV contributes to spontaneous mutagenesis at coding sequences but not microsatellite alleles**. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 2007, **619**(1-2):93-103.
34. Otto SP: **Detecting the form of selection from DNA sequence data**. *Trends in Genetics* 2000, **16**(12):526-529.
35. Cupples CG, Cabrera M, Cruz C, Miller JH: **A Set of lacZ Mutations in Escherichia coli That Allow Rapid Detection of Specific Frameshift Mutations**. *Genetics* 1990, **125**(2):275-280.
36. Levy DD, Cebula TA: **Fidelity of replication of repetitive DNA in mutS and repair proficient Escherichia coli**. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis* 2001, **474**(1-2):1-14.

Figure Legends

Figure 1: Mononucleotide repeats are less frequent in coding regions than in intergenic regions in the genome of *M. tuberculosis*. The figure shows the ratio between total number of observed and expected mononucleotide repeats in coding (filled circles and solid line) and intergenic regions (open circles and dashed line). The expected values have been calculated under the null model that nucleotides are randomly distributed in each of these two compartments, and occur at the frequencies observed in each compartment. For repeats of length three to six nucleotides, the under-representation of mononucleotide repeats is significantly stronger in coding regions than in intergenic regions (at $p < 0.0001$, chi square test).

Figure 2: Proteins of *M. tuberculosis* are encoded in a way that minimizes the emergence of mononucleotide repeats. The lines depict the ratio between the observed and expected number of mononucleotide repeats (summed over all genes in the genome) as a function of their length. The expected numbers were calculated with a null-model that conserved the amino acid sequence and the gene-specific codon frequencies. The areas comprise 95% of the data from the randomized genomes. For repeats of three (for A, C and G) or six (for T) nucleotides and longer, the lines lie below this area, indicating that such repeats are significantly under-represented.

Figure 3: The under-representation of long mononucleotide repeats in coding regions of *M. tuberculosis* is a consequence of a context-dependent codon choice. Codons consisting of three identical nucleotides ('homogeneous codons') are avoided at positions followed by one or more nucleotides of the same type. The under-representation increases with increasing number of identical nucleotides following.

Figure 4: Context-dependent codon choice leads to a strong decrease in the estimated frame-shift mutation rate in mononucleotide repeats in protein-coding genes in *M. tuberculosis* (upper area, light colors), and to a weaker decrease in *E. coli* (lower area, dark colors). We estimated the frame-shift mutation rate in the real genomes, and in randomized genomes without context-dependent codon choice. The ratio between these two values is a measure for the degree by which context-dependent codon choice stabilizes the genome, and is displayed on the z-axis. This ratio was calculated for different combinations of two biological parameters – the increase in the frame-shift mutation rate per nucleotide added to a mononucleotide repeat (parameter c, x-axis), and the minimal length for a mononucleotide repeat to exhibit a significant frame-shift mutation rate (parameter i, y-axis; see Methods). Published data suggest that in MMR-deficient bacteria c is close to 3 and i is about 4. For this combination, context-dependent codon choice leads to a decrease in the frame-shift mutation rate in mononucleotide repeats of a factor of about 20 in *M. tuberculosis* and of a factor of 3 in *E. coli*.

Additional Files

Additional File 1: Tables with observed and expected numbers of mono-, di- and tetranucleotide repeats in protein-coding genes in the genome of *M. tuberculosis.*

Additional File 2: Tables with estimated frame-shift mutation rates in mononucleotide repeats in protein coding genes of *M. tuberculosis* and *E. coli*. Estimates are given for the real genomes as well as for randomized genomes that preserve the amino acid sequence and the gene-specific codon frequencies.

Figure 1

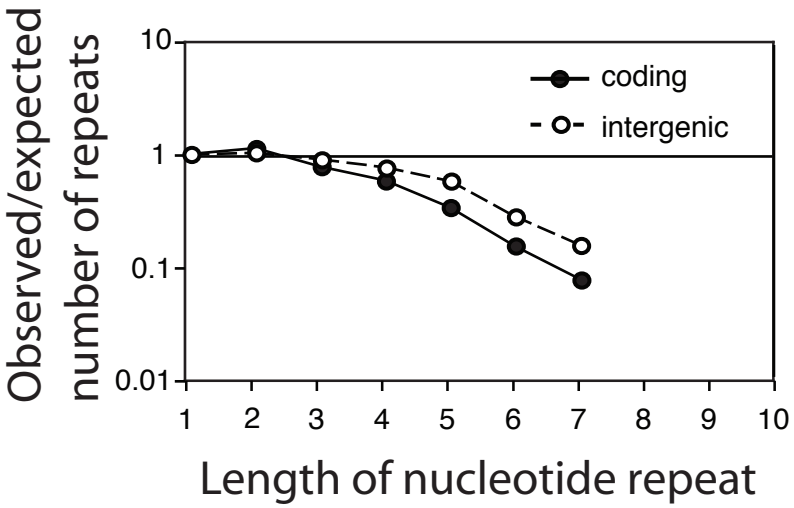


Figure 2

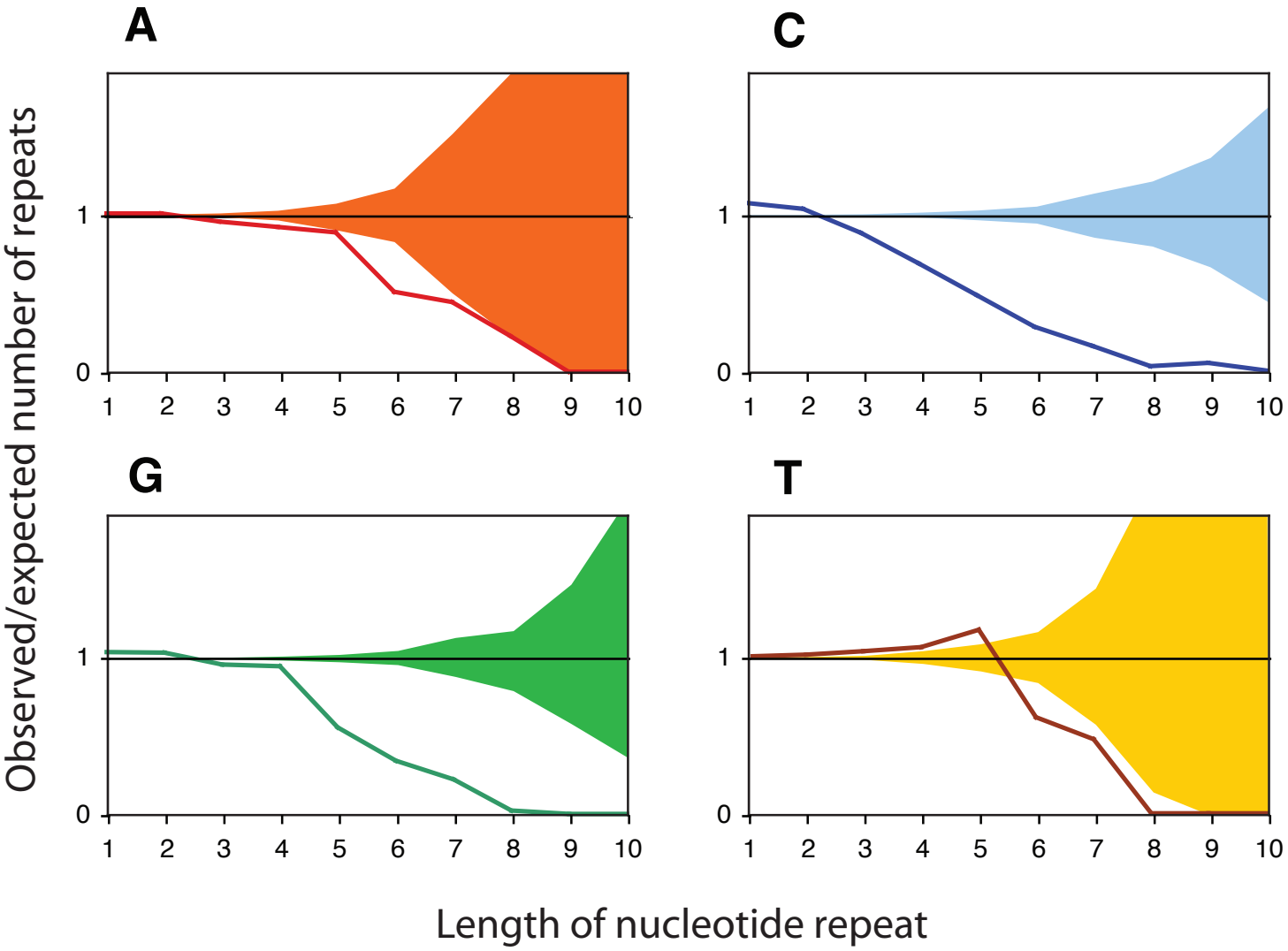
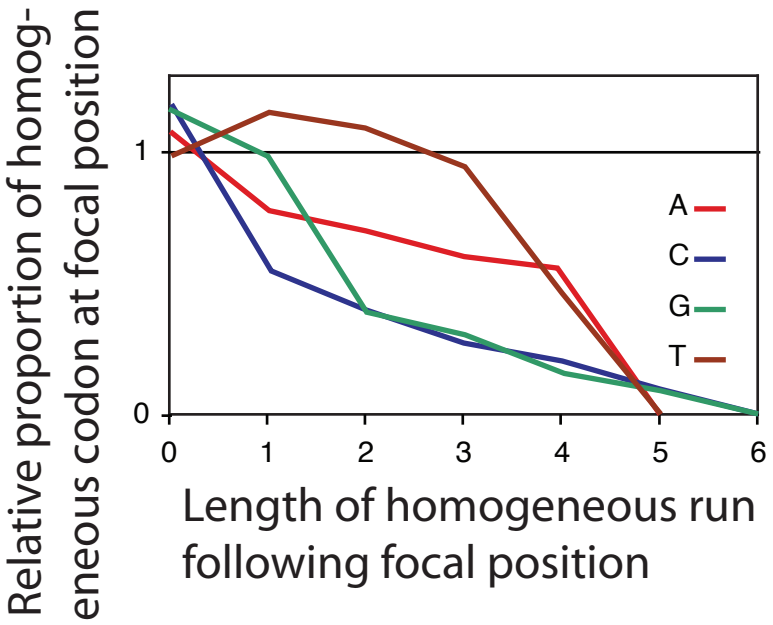

A
C
G
T
Observed/expected number of repeats
0
1
1 2 3 4 5 6 7 8 9 10
Length of nucleotide repeat

Figure 3


Relative proportion of homog-
eneous codon at focal position
1
0
0
1
2
3
4
5
6
A
C
G
T
Length of homogeneous run
following focal position

10000
1000
100
10
1
Decrease in estimated frame-shift mutation rate through context-dependent codon choice
6
5
4
Minimal length (parameter i)
2
2.5
3
3.5
4
Increase in the frame-shift mutation rate per nucleotide added (parameter c)

Additional files provided with this submission:

Additional file 1: additional file 1.pdf, 186K
http://www.biomedcentral.com/imedia/2021514890178502/supp1.pdf
Additional file 2: additional file 2.pdf, 52K
http://www.biomedcentral.com/imedia/8672393317850217/supp2.pdf